

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



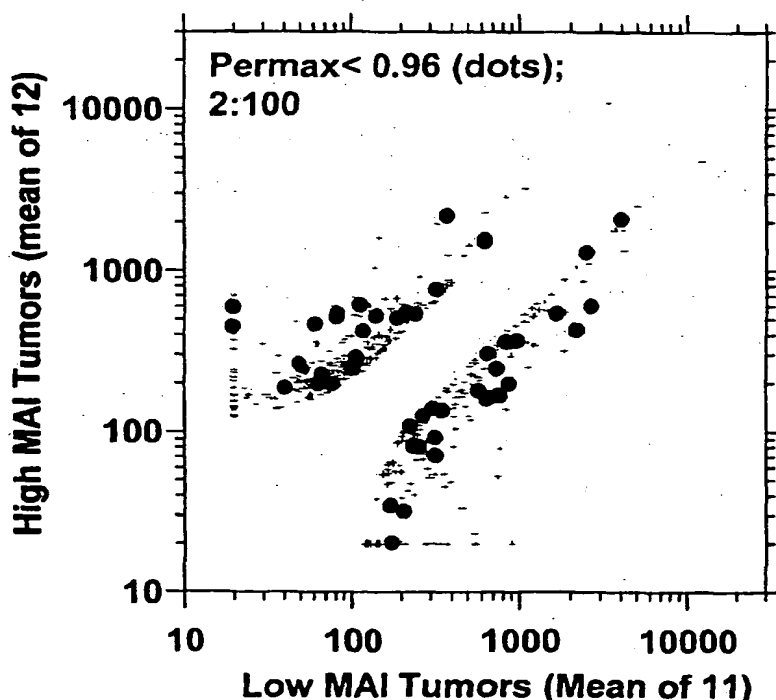
(43) International Publication Date
7 February 2002 (07.02.2002)

PCT

(10) International Publication Number
WO 02/10436 A2

- (51) International Patent Classification⁷: **C12Q 1/00** (72) Inventor: **MUTTER, George, L.**; 230 Middlesex Road, Chestnut Hill, MA 02467-1841 (US).
- (21) International Application Number: **PCT/US01/23642** (74) Agent: **VAN AMSTERDAM, John, R.**; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210 (US).
- (22) International Filing Date: **27 July 2001 (27.07.2001)** (81) Designated States (*national*): **CA, JP.**
- (25) Filing Language: **English** (84) Designated States (*regional*): **European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).**
- (26) Publication Language: **English**
- (30) Priority Data: **60/222,093 28 July 2000 (28.07.2000) US**
- (71) Applicant: **THE BRIGHAM AND WOMEN'S HOSPITAL, INC.** [US/US]; 75 Francis Street, Boston, MA 02115 (US). Published: **— without international search report and to be republished upon receipt of that report**
- (71) Applicant and (72) Inventor: **BAAK, Jan** [NL/NO]; Risajev 66, N-4056 Tananger (NO). *For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: **PROGNOSTIC CLASSIFICATION OF BREAST CANCER**



(57) Abstract: The invention provides particular sets of genes that are expressed differentially in tumors characterized as high MAI or low MAI tumors. These sets of genes can be used to discriminate between high and low MAI tumors. Diagnostic assays for classification of tumors, prediction of tumor outcome, selecting and monitoring treatment regimens and monitoring tumor progression/regression are also provided.

BEST AVAILABLE COPY

PROGNOSTIC CLASSIFICATION OF BREAST CANCER

Field of the Invention

The invention relates to nucleic acid microarray markers for cancer, particularly for breast cancer. The invention also relates to methods for diagnosing cancer as well as optimizing cancer treatment strategies.

Background of the Invention

Breast cancer is a malignant proliferation of epithelial cells lining the ducts or lobules of the breast (Harrison's Principles of Internal Medicine 1998). Although much progress has been made toward understanding the biological basis of cancer and in its diagnosis and treatment, it is still one of the leading causes of death in the United States. Inherent difficulties in the diagnosis and treatment of cancer include among other things, the existence of many different subgroups of cancer and the concomitant variation in appropriate treatment strategies to maximize the likelihood of positive patient outcome.

The traditional method of breast cancer diagnosis and staging is through the use of biopsy examination. Once a diagnosis is made, the options for treating breast cancer are assessed with respect to the needs of the patient. These options traditionally include surgical intervention, chemotherapy, radiotherapy, and adjuvant systemic therapies. Surgical therapy may be lumpectomy or more extensive mastectomy. Adjuvants may include but are not limited to chemotherapy, radiotherapy, and endocrine therapies such as castration; administration of LHRH agonists, antiestrogens, such as tamoxifen, high-dose progestogens; adrenalectomy; and/or aromatase inhibitors (Harrison's Principles of Internal Medicine 1998).

Of key importance in the treatment of breast cancer is the selection and implementation of an appropriate combination of therapeutic approaches. For example, depending on a breast cancer patient's prognosis, therapy may include surgical intervention in combination with adjuvant therapy or it may only include surgical intervention. In addition, for some patients pretreatment with chemotherapy or radiotherapy is utilized prior to surgical intervention, but in other patients adjuvant therapies are used following surgical intervention.

It is difficult to predict from standard clinical and pathologic features the clinical course of early stage breast cancer, particularly lymph node-negative tumors in

- 2 -

premenopausal patients. Current practice in the United States is to offer systemic chemotherapy to most of these women. Because the majority of these women would have good outcome even without chemotherapy, the rate of "over-treatment" is high.

Chemotherapy itself carries a 1% mortality rate. Therefore, unnecessary deaths could be avoided if it were possible to subdivide these patients into high and low risk subgroups, and only undertake adjunctive treatment for those judged to be high risk.

Selection of a suitable treatment regimen for breast cancer is based on the subgroup of cancer. Current strategies used to make therapeutic decisions in the management of patients with breast cancer are based on several factors including hormone receptor status, her-2/neu staining, flow cytometry, and the mitotic activity index (MAI). The MAI is a widely utilized predictor of outcome in cancers, particularly in invasive breast cancer. The definition of the MAI is "the total number of mitoses counted in 10 consecutive high-power fields (objective, x40; numeric aperture, .75; field diameter, 450 microns), in the most cellular area at the periphery of the tumor, with the subjectively highest mitotic activity" (Jannink et al., 1995).

For the procedure, hematoxylin-eosin stained sections of breast cancer tumor are assessed for the total number of mitotic figures in ten consecutive high-power fields and based on these numbers the breast cancer is assigned to either good outcome (MAI<10) or poor outcome (MAI>10). MAI classification correlates to standard parameters such as death, recurrence, and metastases, which are known to those of ordinary skill in the art to predict clinical outcome.

Determination of appropriate treatment for an individual cancer patient is complex with a wide variety of treatments and possible treatment combinations. For example, chemotherapy is a common method of cancer treatment, with more than 50 different chemotherapeutic agents available. These therapeutic agents can be used in a wide range of dosages both singly and in combinational therapies with other chemotherapeutic agents, surgery, and/or radiotherapy.

The available methods for designing strategies for treating breast cancer patients are complex, time consuming, and inexact. The wide range of cancer subgroups and variations in disease progression limit the predictive ability of the healthcare professional. In addition, continuing development of novel treatment strategies and therapeutics will result in the addition of more variables to the already complex decision-making process involving matching the cancer patient with a treatment regimen that is appropriate and optimized for the cancer stage, extent of infiltration, tumor growth rate, and other factors central to the

individual patient's prognosis. Because of the critical importance of selecting appropriate treatment regimens for breast cancer patients, the development of guidelines for treatment selection is of key interest to those in the medical community and their patients. Thus, there presently is a need for objective, reproducible, and sensitive methods for predicting breast cancer patient outcome and selecting optimal treatment regimens.

Summary of the Invention

It now has been discovered that particular sets of genes are expressed differentially in tumors characterized as high MAI or low MAI tumors. These sets of genes can be used to discriminate between high and low MAI tumors. Accordingly, diagnostic assays for classification of tumors, prediction of tumor outcome, selecting and monitoring treatment regimens and monitoring tumor progression/regression can now be based on the expression of sets of genes.

According to one aspect of the invention, methods for diagnosing breast cancer in a subject suspected of having breast cancer are provided. The methods include obtaining from the subject a breast tissue sample and determining the expression of a set of nucleic acid molecules or expression products thereof in the breast tissue sample. The set of nucleic acid molecules includes at least two nucleic acid molecules selected from the group consisting of SEQ ID NOs:1-51. In preferred embodiments, the breast tissue sample suspected of being cancerous.

In some embodiments the set of nucleic acid molecules includes more than 2 and up to all of the nucleic acid molecules set forth as SEQ ID NOs:1-51, and any number of nucleic acid sequences between these two numbers. For example, in certain embodiments the set includes at least 3, 4, 5, 10, 15, 20, 30, 40 or more nucleic acid molecules of the nucleic acid molecules set forth as SEQ ID NOs:1-51.

In other embodiments, the method further includes determining the expression of the set of nucleic acid molecules or expression products thereof in a non-cancerous breast tissue sample, and comparing the expression of the set of nucleic acid molecules or expression products thereof in the breast tissue sample suspected of being cancerous and the non-cancerous breast tissue sample.

According to another aspect of the invention, methods for identifying a set of nucleic acid markers or expression products thereof are provided. The methods are effective for determining the prognosis of cancer. The methods include obtaining a plurality of tumor

tissue samples from a plurality of subjects afflicted with cancer, classifying the plurality of tumor tissue samples according to mitotic activity index (MAI) into high MAI and low MAI groups and determining differences in the expression of a plurality of nucleic acid molecules or expression products thereof in the tumor tissue samples. The methods further include
5 selecting as a set of nucleic acid markers the nucleic acid molecules or expression products thereof which are differentially expressed in the high MAI and the low MAI groups. The set of nucleic acid markers or expression products thereof effective for determining poor prognosis of cancer includes one or more nucleic acid molecules or expression products thereof which are preferentially expressed in high MAI tumor tissue samples, and wherein the
10 set of nucleic acid markers or expression products thereof effective for determining good prognosis of cancer comprises one or more nucleic acid molecules or expression products thereof which are preferentially expressed in low MAI tumor tissue samples. In preferred embodiments, the cancer is breast cancer.

According to still another aspect of the invention, methods for selecting a course of
15 treatment of a subject having or suspected of having cancer are provided. The methods include obtaining from the subject a tissue sample suspected of being cancerous, determining the expression of a set of nucleic acid markers or expression products thereof which are differentially expressed in high MAI tumor tissue samples to determine the MAI of the tissue sample of the subject, and selecting a course of treatment appropriate to the cancer of the
20 subject.

In preferred embodiments the cancer is breast cancer, and in some of these embodiments the methods include determining the expression of a set of nucleic acid markers that are differentially expressed in low MAI breast tumor tissue samples.

According to yet another aspect of the invention, methods for evaluating treatment of
25 cancer are provided. The methods include obtaining a first determination of the expression of a set of nucleic acid molecules or expression products thereof, which are differentially expressed in high MAI tumor tissue samples to determine the MAI of the tissue sample from a subject undergoing treatment for cancer, and obtaining a second determination of the expression of a set of nucleic acid molecules or expression products thereof, which are
30 differentially expressed in high MAI tumor tissue samples to determine the MAI of the second tissue sample from the subject after obtaining the first determination. The methods also include comparing the first determination of expression to the second determination of expression as an indication of evaluation of the treatment.

In preferred embodiments the cancer is breast cancer, and in some of these embodiments the methods include determining the expression of a set of nucleic acid markers that are differentially expressed in low MAI breast tumor tissue samples.

The invention in another aspect provides solid-phase nucleic acid molecule arrays.

5 The arrays have a cancer gene marker set that consists essentially of at least two and as many as all of the nucleic acid molecules set forth as SEQ ID NOs:1-51 fixed to a solid substrate. The set of nucleic acid markers can include any number of nucleic acid sequences between these two numbers, selected from SEQ ID NOs:1-51. For example, in certain embodiments the set includes at least 3, 4, 5, 10, 15, 20, 30, 40 or more nucleic acid molecules of the
10 nucleic acid molecules set forth as SEQ ID NOs:1-51. In some embodiments, the solid-phase nucleic acid molecule array also includes at least one control nucleic acid molecule.

In certain embodiments, the solid substrate includes a material selected from the group consisting of glass, silica, aluminosilicates, borosilicates, metal oxides such as alumina and nickel oxide, various clays, nitrocellulose, or nylon. Preferably the substrate is glass.

15 In other embodiments, the nucleic acid molecules are fixed to the solid substrate by covalent bonding.

According to yet another aspect of the invention, protein microarrays are provided.

The protein microarrays include antibodies or antigen-binding fragments thereof, that specifically bind at least two different polypeptides selected from the group consisting of
20 SEQ ID NOs:52-102, fixed to a solid substrate. In some embodiments, the microarray comprises antibodies or antigen-binding fragments thereof, that bind specifically to least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 or 51 different polypeptides selected from the group consisting of SEQ ID NOs:52-102. In certain
25 embodiments, the microarray also includes an antibody or antigen-binding fragment thereof, that binds specifically to a cancer-associated polypeptide other than those selected from the group consisting of SEQ ID NOs:52-102, preferably a breast cancer associated polypeptide. In some embodiments, the protein microarray also includes at least one control polypeptide molecule. In further embodiments, the antibodies are monoclonal or polyclonal antibodies.
30 In other embodiments, the antibodies are chimeric, human, or humanized antibodies. In some embodiments, the antibodies are single chain antibodies. In still other embodiments, the antigen-binding fragments are F(ab')₂, Fab, Fd, or Fv fragments.

- 6 -

In a further aspect of the invention, methods for identifying lead compounds for a pharmacological agent useful in the treatment of breast cancer are provided. The methods include contacting a breast cancer cell or tissue with a candidate pharmacological agent, and determining the expression of a set of nucleic acid molecules in the breast cancer cell or
5 tissue sample under conditions which, in the absence of the candidate pharmacological agent, permit a first amount of expression of the set of nucleic acid molecules. The set of nucleic acid molecules includes at least two and as many as all of the nucleic acid molecules set forth as SEQ ID NOs:1-51. The methods also include detecting a test amount of the expression of the set of nucleic acid molecules, wherein a decrease in the test amount of expression in the
10 presence of the candidate pharmacological agent relative to the first amount of expression indicates that the candidate pharmacological agent is a lead compound for a pharmacological agent which is useful in the treatment of breast cancer. In preferred embodiments, the set of nucleic acid molecules is differentially expressed in high MAI breast tumor tissue samples.

In some embodiments of any of the foregoing methods and products, the differences
15 in the expression of a the nucleic acid molecules are determined by nucleic acid hybridization or nucleic acid amplification methods. Preferably the nucleic acid hybridization is performed using a solid-phase nucleic acid molecule array. In other embodiments, the differences in the expression of the nucleic acid molecules are determined by protein expression analysis, preferably SELDI mass spectroscopy.

20 These and other aspects of the invention will be described in greater detail below.

Brief Description of the Drawings

Figure 1 is a scatterplot of gene expression level in low risk (x axis) and high risk (y axis) breast cancers. 422 genes whose mean expression between groups differs at least 2-fold
25 and by 100 expression units are shown as small crosses. The top 51 t-test ranked genes with Permax 0.96 are indicated as solid circles, and appear in Table 1.

Detailed Description of the Invention

The invention described herein relates to the identification of a set of genes expressed
30 in breast cancer tissue that are predictive of the clinical outcome of the cancer. Changes in cell phenotype in cancer are often the result of one or more changes in the genome expression of the cell. Some genes are expressed in tumor cells, and not in normal cells. In addition, different genes are expressed in different subgroups of breast cancers, which have different

prognoses and require different treatment regimens to optimize patient outcome. The differential expression of breast cancer genes can be examined by the assessment of nucleic acid or protein expression in the breast cancer tissue.

The genes were identified by screening nucleic acid molecules isolated from various breast cancer samples for expression of the genes present on a high-density nucleic acid microarray. The breast cancer samples were categorized with respect to their mitotic activity index (MAI) and the MAI was correlated to gene expression to identify those genes differentially expressed between low and high-MAI breast cancer tissue. The MAI has been shown to correlate with the outcome of the cancer as defined by tumor metastasis, tumor recurrence or mortality. Accordingly the genes identified permit, *inter alia*, rapid screening of cancer samples by nucleic acid microarray hybridization or protein expression technology to determine the expression of the specific genes and thereby to predict the outcome of the cancer. Such screening is beneficial, for example, in selecting the course of treatment to provide to the cancer patient, and to monitor the efficacy of a treatment.

The invention differs from traditional breast cancer diagnostic and classification techniques including MAI, hormone receptor expression and her-2/neu expression, with respect to the speed, simplicity, and reproducibility of the cancer diagnostic assay. The invention also presents targets for drug development because it identifies genes that are differentially expressed in poor outcome breast tumors, which can be utilized in the development of drugs to treat such tumors, e.g., by reducing expression of the genes or reducing activity of proteins encoded by the genes.

The invention moves beyond the use of the MAI and simplifies prognosis determination by providing an identified set of genes whose expression in breast cancers predicts poor clinical outcome as defined by tumor metastasis, recurrence, or death. In the invention, the MAI was used in conjunction with RNA expression phenotyping performed using high density microarrays generated from quantitative expression data on over 5000 (estimated 5800) genes, which have been analyzed to identify 51 specific probe sets (genes) with divergent expression between MAI groups. The expression gene set has multifold uses including, but not limited to, the following examples. The expression gene set may be used as a prognostic tool for breast cancer patients, to make possible more finely tuned diagnosis of breast cancer and allow healthcare professionals to tailor treatment to individual patients' needs. The invention can also assess the efficacy of breast cancer treatment by determining progression or regression of breast cancer in patients before, during, and after breast cancer

treatment. Another utility of the expression gene set is in the biotechnology and pharmaceutical industries' research on disease pathway discovery for therapeutic targeting. The invention can identify alterations in gene expression in breast cancer and can also be used to uncover and test candidate pharmaceutical agents to treat breast cancer.

5 Although the invention is described primarily with respect to breast cancer, one of ordinary skill in the art will appreciate that the invention also is useful for diagnosis and prognosis determination of cancers that can be classified into subgroups for prognosis of the cancer based on MAI. For example, MAI has been used successfully in the classification of malignant melanoma, ovarian cancer, bladder cancer, and prostatic adenocarcinoma. Thus,
10 the methods and products of the invention also are applicable to non-breast cancers that can be classified by MAI.

 The invention may also encompass cancers other than breast cancer, including but not limited to: biliary tract cancer; bladder cancer; brain cancer including glioblastomas and medulloblastomas; cervical cancer; choriocarcinoma; colon cancer; endometrial cancer;
15 esophageal cancer; gastric cancer; hematological neoplasms including acute lymphocytic and myelogenous leukemia; multiple myeloma; AIDS-associated leukemias and adult T-cell leukemia lymphoma; intraepithelial neoplasms including Bowen's disease and Paget's disease; liver cancer; lung cancer; lymphomas including Hodgkin's disease and lymphocytic lymphomas; neuroblastomas; oral cancer including squamous cell carcinoma; ovarian cancer
20 including those arising from epithelial cells, stromal cells, germ cells and mesenchymal cells; pancreatic cancer; prostate cancer; rectal cancer; sarcomas including leiomyosarcoma, rhabdomyosarcoma, liposarcoma, fibrosarcoma, and osteosarcoma; skin cancer including melanoma, Kaposi's sarcoma, basocellular cancer, and squamous cell cancer; testicular cancer including germinal tumors such as seminoma, non-seminoma (teratomas,
25 choriocarcinomas), stromal tumors, and germ cell tumors; thyroid cancer including thyroid adenocarcinoma and medullar carcinoma; and renal cancer including adenocarcinoma and Wilms tumor.

 As used herein, a subject is a human, non-human primate, cow, horse, pig, sheep, goat, dog, cat or rodent. In all embodiments human subjects are preferred. Preferably the
30 subject is a human either suspected of having breast cancer, or having been diagnosed with breast cancer. In a preferred embodiment of the invention the cancer is pre-menopausal, lymph node-negative breast cancer. Methods for identifying subjects suspected of having breast cancer may include manual examination, biopsy, subject's family medical history,

subject's medical history, or a number of imaging technologies such as mammography, magnetic resonance imaging, magnetic resonance spectroscopy, or positron emission tomography. Diagnostic methods for breast cancer and the clinical delineation of breast cancer diagnoses are well-known to those of skill in the medical arts.

5 As used herein, breast tissue sample is tissue obtained from a breast tissue biopsy using methods well-known to those of ordinary skill in the related medical arts. The phrase "suspected of being cancerous" as used herein means a breast cancer tissue sample believed by one of ordinary skill in the medical arts to contain cancerous cells. Methods for obtaining the sample from the biopsy include gross apportioning of a mass, microdissection, laser-
10 based microdissection, or other art-known cell-separation methods.

Because of the variability of the cell types in diseased-tissue biopsy material, and the variability in sensitivity of the diagnostic methods used, the sample size required for analysis may range from 1, 10, 50, 100, 200, 300, 500, 1000, 5000, 10,000, to 50,000 or more cells. The appropriate sample size may be determined based on the cellular composition and
15 condition of the biopsy and the standard preparative steps for this determination and subsequent isolation of the nucleic acid for use in the invention are well known to one of ordinary skill in the art. An example of this, although not intended to be limiting, is that in some instances a sample from the biopsy may be sufficient for assessment of RNA expression without amplification, but in other instances the lack of suitable cells in a small
20 biopsy region may require use of RNA conversion and/or amplification methods or other methods to enhance resolution of the nucleic acid molecules. Such methods, which allow use of limited biopsy materials, are well known to those of ordinary skill in the art and include, but are not limited to: direct RNA amplification, reverse transcription of RNA to cDNA, amplification of cDNA, or the generation of radio-labeled nucleic acids.

25 As used herein, the phrase "determining the expression of a set of nucleic acid molecules in the breast tissue" means identifying RNA transcripts in the tissue sample by analysis of nucleic acid or protein expression in the tissue sample. As used herein, "set" refers to a group of nucleic acid molecules that include 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39,
30 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or 51 different nucleic acid sequences from the group of nucleic acid sequences numbered 1 through 51 in Table 1 (SEQ ID Nos: 1-51).

The expression of the set of nucleic acid molecules in the sample from the breast cancer patient can be compared to the expression of the set of nucleic acid molecules in a

sample of breast tissue that is non-cancerous. As used herein, non-cancerous breast tissue means tissue determined by one of ordinary skill in the medical art to have no evidence of breast cancer based on standard diagnostic methods including, but not limited to, histologic staining and microscopic analysis.

5 Nucleic acid markers for cancer are nucleic acid molecules that by their presence or absence indicate the presence of absence of breast cancer. In tissue, certain nucleic acid molecules are expressed at different levels depending on whether tissue is non-cancerous or cancerous. In cancerous tissue, nucleic acid molecule expression may be correlated with MAI prognostic analysis. As described herein, breast cancer nucleic acid markers were
10 identified by evaluating the nucleic acid molecules present in breast tumor tissue samples and comparing expression levels of the nucleic acid molecules with MAI levels determined for the tissues. An aspect of the invention is that different nucleic acid molecules are expressed in breast cancers with different MAI levels (i.e., high MAI versus low MAI) and these expression variations are identifiable by nucleic acid expression analysis, such as microarray
15 analysis or protein expression analysis. Some nucleic acids are more likely to be, in other words, are preferentially expressed in cancers with high MAI levels and other nucleic acids are preferentially expressed in cancers with low MAI levels. According to the invention, the correlation between the preferential expression of nucleic acid markers and MAI classification allows expression of nucleic acid markers to be used to directly categorize
20 breast cancers as low MAI or high MAI. Thus, nucleic acid expression-based categorization of breast cancer (by measurement of nucleic acid or protein expression) as low or high MAI may be used by one of ordinary skill in the medical arts to select an appropriate treatment regimen based on a patient's specific breast cancer prognosis.

Hybridization methods for nucleic acids are well known to those of ordinary skill in
25 the art (see, e.g. *Molecular Cloning: A Laboratory Manual*, J. Sambrook, et al., eds., Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989, or *Current Protocols in Molecular Biology*, F.M. Ausubel, et al., eds., John Wiley & Sons, Inc., New York). The nucleic acid molecules from a breast cancer tissue sample hybridize under stringent conditions to nucleic acid markers expressed in breast cancer. In one embodiment
30 the markers are sets of two or more of the nucleic acid molecules as set forth in SEQ ID NOs: 1 through 51.

The breast cancer nucleic acid markers disclosed herein are known genes and fragments thereof. It may be desirable to identify variants of those genes, such as allelic

variants or single nucleotide polymorphisms (SNPs) in tissues. Accordingly, methods for identifying breast cancer nucleic acid markers, including variants of the disclosed full-length cDNAs, genomic DNAs, and SNPs are also included in the invention. The methods include contacting a nucleic acid sample (such as a cDNA library, genomic library, genomic DNA isolate, etc.) with a nucleic acid probe or primer derived from one of SEQ ID NOs:1 through 51. The nucleic acid sample and the probe or primer hybridize to complementary nucleotide sequences of nucleic acids in the sample, if any are present, allowing detection of nucleic acids related to SEQ ID NOs: 1-51. Preferably the probe or primer is detectably labeled. The specific conditions, reagents, and the like can be selected by one of ordinary skill in the art to selectively identify nucleic acids related to sets of two or more of SEQ ID NOs:1 through 51. The isolated nucleic acid molecule can be sequenced according to standard procedures.

In addition to native nucleic acid markers (SEQ ID NOs:1-51), the invention also includes degenerate nucleic acids that include alternative codons to those present in the native materials. For example, serine residues are encoded by the codons TCA, AGT, TCC, TCG, TCT, and AGC. Each of the six codons is equivalent for the purposes of encoding a serine residue. Similarly, nucleotide sequence triplets that encode other amino acid residues include, but are not limited to: CCA, CCC, CCG, and CCT (proline codons); CGA, CGC, CGG, CGT, AGA, and AGG (arginine codons); ACA, ACC, ACG, and ACT (threonine codons); AAC and AAT (asparagine codons); and ATA, ATC, and ATT (isoleucine codons). Other amino acid residues may be encoded similarly by multiple nucleotide sequences. Thus, the invention embraces degenerate nucleic acids that differ from the biologically isolated nucleic acids in codon sequence due to the degeneracy of the genetic code.

The invention also provides modified nucleic acid molecules, which include additions, substitutions, and deletions of one or more nucleotides such as the allelic variants and SNPs described above. In preferred embodiments, these modified nucleic acid molecules and/or the polypeptides they encode retain at least one activity or function of the unmodified nucleic acid molecule and/or the polypeptides, such as hybridization, antibody binding, etc. In certain embodiments, the modified nucleic acid molecules encode modified polypeptides, preferably polypeptides having conservative amino acid substitutions. As used herein, a "conservative amino acid substitution" refers to an amino acid substitution which does not alter the relative charge or size characteristics of the protein in which the amino acid substitution is made. Conservative substitutions of amino acids include substitutions made amongst amino acids within the following groups: (a) M, I, L, V; (b) F, Y, W; (c) K, R, H;

(d) A, G; (e) S, T; (f) Q, N; and (g) E, D. The modified nucleic acid molecules are structurally related to the unmodified nucleic acid molecules and in preferred embodiments are sufficiently structurally related to the unmodified nucleic acid molecules so that the modified and unmodified nucleic acid molecules hybridize under stringent conditions known to one of skill in the art.

For example, modified nucleic acid molecules that encode polypeptides having single amino acid changes can be prepared for use in the methods and products disclosed herein. Each of these nucleic acid molecules can have one, two, or three nucleotide substitutions exclusive of nucleotide changes corresponding to the degeneracy of the genetic code as described herein. Likewise, modified nucleic acid molecules that encode polypeptides having two amino acid changes can be prepared, which have, e.g., 2-6 nucleotide changes. Numerous modified nucleic acid molecules like these will be readily envisioned by one of skill in the art, including for example, substitutions of nucleotides in codons encoding amino acids 2 and 3, 2 and 4, 2 and 5, 2 and 6, and so on. In the foregoing example, each combination of two amino acids is included in the set of modified nucleic acid molecules, as well as all nucleotide substitutions which code for the amino acid substitutions. Additional nucleic acid molecules that encode polypeptides having additional substitutions (i.e., 3 or more), additions or deletions [e.g., by introduction of a stop codon or a splice site(s)] also can be prepared and are embraced by the invention as readily envisioned by one of ordinary skill in the art. Any of the foregoing nucleic acids can be tested by routine experimentation for retention of structural relation to or activity similar to the nucleic acids disclosed herein.

In the invention, standard hybridization techniques of microarray technology are utilized to assess patterns of nucleic acid expression and identify nucleic acid marker expression. Microarray technology, which is also known by other names including: DNA chip technology, gene chip technology, and solid-phase nucleic acid array technology, is well known to those of ordinary skill in the art and is based on, but not limited to, obtaining an array of identified nucleic acid probes on a fixed substrate, labeling target molecules with reporter molecules (e.g., radioactive, chemiluminescent, or fluorescent tags such as fluorescein, Cy3-dUTP, or Cy5-dUTP), hybridizing target nucleic acids to the probes, and evaluating target-probe hybridization. A probe with a nucleic acid sequence that perfectly matches the target sequence will, in general, result in detection of a stronger reporter-molecule signal than will probes with less perfect matches. Many components and techniques utilized in nucleic acid microarray technology are presented in *The Chipping*

Forecast, Nature Genetics, Vol.21, Jan 1999, the entire contents of which is incorporated by reference herein.

According to the present invention, microarray substrates may include but are not limited to glass, silica, aluminosilicates, borosilicates, metal oxides such as alumina and nickel oxide, various clays, nitrocellulose, or nylon. In all embodiments a glass substrate is preferred. According to the invention, probes are selected from the group of nucleic acids including, but not limited to: DNA, genomic DNA, cDNA, and oligonucleotides; and may be natural or synthetic. Oligonucleotide probes preferably are 20 to 25-mer oligonucleotides and DNA/cDNA probes preferably are 500 to 5000 bases in length, although other lengths may be used. Appropriate probe length may be determined by one of ordinary skill in the art by following art-known procedures. In one embodiment, preferred probes are sets of two or more of the nucleic acid molecules set forth as SEQ ID NO: 1 through 51 (see also Table 1). Probes may be purified to remove contaminants using standard methods known to those of ordinary skill in the art such as gel filtration or precipitation.

In one embodiment, the microarray substrate may be coated with a compound to enhance synthesis of the probe on the substrate. Such compounds include, but are not limited to, oligoethylene glycols. In another embodiment, coupling agents or groups on the substrate can be used to covalently link the first nucleotide or oligonucleotide to the substrate. These agents or groups may include, but are not limited to: amino, hydroxy, bromo, and carboxy groups. These reactive groups are preferably attached to the substrate through a hydrocarbyl radical such as an alkylene or phenylene divalent radical, one valence position occupied by the chain bonding and the remaining attached to the reactive groups. These hydrocarbyl groups may contain up to about ten carbon atoms, preferably up to about six carbon atoms. Alkylene radicals are usually preferred containing two to four carbon atoms in the principal chain. These and additional details of the process are disclosed, for example, in U.S. Patent 4,458,066, which is incorporated by reference in its entirety.

In one embodiment, probes are synthesized directly on the substrate in a predetermined grid pattern using methods such as light-directed chemical synthesis, photochemical deprotection, or delivery of nucleotide precursors to the substrate and subsequent probe production.

In another embodiment, the substrate may be coated with a compound to enhance binding of the probe to the substrate. Such compounds include, but are not limited to: polylysine, amino silanes, amino-reactive silanes (Chipping Forecast, 1999) or chromium

(Gwynne and Page, 2000). In this embodiment, presynthesized probes are applied to the substrate in a precise, predetermined volume and grid pattern, utilizing a computer-controlled robot to apply probe to the substrate in a contact-printing manner or in a non-contact manner such as ink jet or piezo-electric delivery. Probes may be covalently linked to the substrate with methods that include, but are not limited to, UV-irradiation. In another embodiment probes are linked to the substrate with heat.

Targets are nucleic acids selected from the group, including but not limited to: DNA, genomic DNA, cDNA, RNA, mRNA and may be natural or synthetic. In all embodiments, nucleic acid molecules from human breast tissue are preferred. The tissue may be obtained from a subject or may be grown in culture (e.g. from a breast cancer cell line).

In embodiments of the invention one or more control nucleic acid molecules are attached to the substrate. Preferably, control nucleic acid molecules allow determination of factors including but not limited to: nucleic acid quality and binding characteristics; reagent quality and effectiveness; hybridization success; and analysis thresholds and success. Control nucleic acids may include but are not limited to expression products of genes such as housekeeping genes or fragments thereof.

To select a set of tumor markers, the expression data generated by, for example, microarray analysis of gene expression, is preferably analyzed to determine which genes in different groups of cancer tissues are significantly differentially expressed. In the methods disclosed herein, the significance of gene expression was determined using Permax computer software, although any standard statistical package that can discriminate significant differences in expression may be used. Permax performs permutation 2-sample t-tests on large arrays of data. For high dimensional vectors of observations, the Permax software computes t-statistics for each attribute, and assesses significance using the permutation distribution of the maximum and minimum overall attributes. The main use is to determine the attributes (genes) that are the most different between two groups (e.g., high MAI tissues versus low MAI tissues), measuring "most different" using the value of the t-statistics, and their significance levels.

In one embodiment of the invention, expression of nucleic acid markers is used to select clinical treatment paradigms for breast cancer. Treatment options, as described herein, may include but are not limited to: chemotherapy, radiotherapy, adjuvant therapy, or any combination of the aforementioned methods. Aspects of treatment that may vary include, but are not limited to: dosages, timing of administration, or duration or therapy; and may or may

not be combined with other treatments, which may also vary in dosage, timing, or duration. Another treatment for breast cancer is surgery, which can be utilized either alone or in combination with any of the aforementioned treatment methods. One of ordinary skill in the medical arts may determine an appropriate treatment paradigm based on evaluation of differential expression of sets of two or more of the nucleic acid targets SEQ ID NOs:1-51. Cancers that express markers that are indicative of a more aggressive cancer or poor prognosis may be treated with more aggressive therapies.

Progression or regression of breast cancer is determined by comparison of two or more different breast cancer tissue samples taken at two or more different times from a subject. For example, progression or regression may be evaluated by assessments of expression of sets of two or more of the nucleic acid targets, including but not limited to SEQ ID NOs:1-51, in a breast cancer tissue sample from a subject before, during, and following treatment for breast cancer.

In another embodiment, novel pharmacological agents useful in the treatment of breast cancer can be identified by assessing variations in the expression of sets of two or more breast cancer nucleic acid markers, from among SEQ ID NOs:1-51, prior to and after contacting breast cancer cells or tissues with candidate pharmacological agents for the treatment of breast cancer. The cells may be grown in culture (e.g. from a breast cancer cell line), or may be obtained from a subject, (e.g. in a clinical trial of candidate pharmaceutical agents to treat breast cancer). Alterations in expression of two or more sets of breast cancer nucleic acid markers, from among SEQ ID NOs:1-51, in breast cancer cells or tissues tested before and after contact with a candidate pharmacological agent to treat breast cancer, indicate progression, regression, or stasis of the breast cancer thereby indicating efficacy of candidate agents and concomitant identification of lead compounds for therapeutic use in breast cancer.

The invention further provides efficient methods of identifying pharmacological agents or lead compounds for agents active at the level of breast cancer cellular function. Generally, the screening methods involve assaying for compounds that beneficially alter breast cancer nucleic acid molecule expression. Such methods are adaptable to automated, high throughput screening of compounds.

The assay mixture comprises a candidate pharmacological agent. Typically, a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a different response to the various concentrations. Typically, one of these concentrations

serves as a negative control, i.e., at zero concentration of agent or at a concentration of agent below the limits of assay detection. Candidate agents encompass numerous chemical classes, although typically they are organic compounds. Preferably, the candidate pharmacological agents are small organic compounds, i.e., those having a molecular weight of more than 50 yet less than about 2500, preferably less than about 1000 and, more preferably, less than about 500. Candidate agents comprise functional chemical groups necessary for structural interactions with polypeptides and/or nucleic acids, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups and more preferably at least three of the functional chemical groups. The candidate agents can comprise cyclic carbon or heterocyclic structure and/or aromatic or polyaromatic structures substituted with one or more of the above-identified functional groups. Candidate agents also can be biomolecules such as peptides, saccharides, fatty acids, sterols, isoprenoids, purines, pyrimidines, derivatives or structural analogs of the above, or combinations thereof and the like. Where the agent is a nucleic acid, the agent typically is a DNA or RNA molecule, although modified nucleic acids as defined herein are also contemplated.

Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides, synthetic organic combinatorial libraries, phage display libraries of random peptides, and the like. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural and synthetically produced libraries and compounds can be readily be modified through conventional chemical, physical, and biochemical means. Further, known pharmacological agents may be subjected to directed or random chemical modifications such as acylation, alkylation, esterification, amidification, etc. to produce structural analogs of the agents.

A variety of other reagents also can be included in the mixture. These include reagents such as salts, buffers, neutral proteins (e.g., albumin), detergents, etc. which may be used to facilitate optimal protein-protein and/or protein-nucleic acid binding. Such a reagent may also reduce non-specific or background interactions of the reaction components. Other reagents that improve the efficiency of the assay such as protease, inhibitors, nuclease inhibitors, antimicrobial agents, and the like may also be used.

The mixture of the foregoing assay materials is incubated under conditions whereby, the anti-breast cancer candidate agent specifically binds the cellular binding target, a portion thereof or analog thereof. The order of addition of components, incubation temperature, time of incubation, and other parameters of the assay may be readily determined. Such experimentation merely involves optimization of the assay parameters, not the fundamental composition of the assay. Incubation temperatures typically are between 4°C and 40°C. Incubation times preferably are minimized to facilitate rapid, high throughput screening, and typically are between 0.1 and 10 hours.

After incubation, the presence or absence of specific binding between the anti-breast cancer candidate agent and one or more binding targets is detected by any convenient method available to the user. For cell-free binding type assays, a separation step is often used to separate bound from unbound components. The separation step may be accomplished in a variety of ways. Conveniently, at least one of the components is immobilized on a solid substrate, from which the unbound components may be easily separated. The solid substrate can be made of a wide variety of materials and in a wide variety of shapes, e.g., microtiter plate, microbead, dipstick, resin particle, etc. The substrate preferably is chosen to maximize signal to noise ratios, primarily to minimize background binding, as well as for ease of separation and cost.

Separation may be effected for example, by removing a bead or dipstick from a reservoir, emptying or diluting a reservoir such as a microtiter plate well, rinsing a bead, particle, chromatographic column or filter with a wash solution or solvent. The separation step preferably includes multiple rinses or washes. For example, when the solid substrate is a microtiter plate, the wells may be washed several times with a washing solution, which typically includes those components of the incubation mixture that do not participate in specific bindings such as salts, buffer, detergent, non-specific protein, etc. Where the solid substrate is a magnetic bead, the beads may be washed one or more times with a washing solution and isolated using a magnet.

Detection may be effected in any convenient way for cell-based assays such as two- or three-hybrid screens. The transcript resulting from a reporter gene transcription assay of the anti-cancer agent binding to a target molecule typically encodes a directly or indirectly detectable product, e.g., β -galactosidase activity, luciferase activity, and the like. For cell-free binding assays, one of the components usually comprises, or is coupled to, a detectable label. A wide variety of labels can be used, such as those that provide direct detection (e.g.,

radioactivity, luminescence, optical or electron density, etc). or indirect detection (e.g., epitope tag such as the FLAG epitope, enzyme tag such as horseshoe peroxidase, etc.). The label may be bound to an anti-cancer agent binding partner, or incorporated into the structure of the binding partner.

5 A variety of methods may be used to detect the label, depending on the nature of the label and other assay components. For example, the label may be detected while bound to the solid substrate or subsequent to separation from the solid substrate. Labels may be directly detected through optical or electron density, radioactive emissions, nonradiative energy transfers, etc. or indirectly detected with antibody conjugates, streptavidin-biotin conjugates,
10 etc. Methods for detecting the labels are well known in the art.

The invention provides breast cancer gene-specific binding agents, methods of identifying and making such agents, and their use in diagnosis, therapy and pharmaceutical development. For example, breast cancer gene-specific pharmacological agents are useful in a variety of diagnostic and therapeutic applications as described herein. In general, the
15 specificity of a breast cancer gene binding to a binding agent is shown by binding equilibrium constants. Targets which are capable of selectively binding a breast cancer gene preferably have binding equilibrium constants of at least about 10^7 M^{-1} , more preferably at least about 10^8 M^{-1} , and most preferably at least about 10^9 M^{-1} . The wide variety of cell based and cell free assays may be used to demonstrate breast cancer gene-specific binding. Cell-based
20 assays include one, two and three hybrid screens, assays in which breast cancer gene-mediated transcription is inhibited or increased, etc. Cell-free assays include breast cancer gene-protein binding assays, immunoassays, etc. Other assays useful for screening agents which bind breast cancer polypeptides include fluorescence resonance energy transfer (FRET), and electrophoretic mobility shift analysis (EMSA).

25 In another aspect of the invention, pre- and post-treatment alterations in expression of two or more sets of breast cancer nucleic acid markers including, but not limited to, SEQ ID NOs:1-51 in breast cancer cells or tissues may be used to assess treatment parameters including, but not limited to: dosage, method of administration, timing of administration, and combination with other treatments as described herein.

30 Candidate pharmacological agents may include antisense oligonucleotides that selectively binds to a breast cancer nucleic acid marker molecule, as identified herein, to reduce the expression of the marker molecules in breast cancer cells and tissues. One of ordinary skill in the art can test of the effects of a reduction of expression of breast cancer

nucleic acid marker sequences *in vivo* or *in vitro*, to determine the efficacy of one or more antisense oligonucleotides.

As used herein, the term "antisense oligonucleotide" or "antisense" describes an oligonucleotide that is an oligoribonucleotide, oligodeoxyribonucleotide, modified
5 oligoribonucleotide, or modified oligodeoxyribonucleotide which hybridizes under physiological conditions to DNA comprising a particular gene or to an mRNA transcript of that gene and, thereby, inhibits the transcription of that gene and/or the translation of that mRNA. The antisense molecules are designed so as to interfere with transcription or translation of a target gene upon hybridization with the target gene or transcript. Those
10 skilled in the art will recognize that the exact length of the antisense oligonucleotide and its degree of complementarity with its target will depend upon the specific target selected, including the sequence of the target and the particular bases which comprise that sequence. It is preferred that the antisense oligonucleotide be constructed and arranged so as to bind selectively with the target under physiological conditions, i.e., to hybridize substantially more
15 to the target sequence than to any other sequence in the target cell under physiological conditions.

Based upon the sequences of breast cancer expressed nucleic acids, or upon allelic or homologous genomic and/or cDNA sequences, one of skill in the art can easily choose and synthesize any of a number of appropriate antisense molecules for use in accordance with the
20 present invention. In order to be sufficiently selective and potent for inhibition, such antisense oligonucleotides should comprise at least 10 and, more preferably, at least 15 consecutive bases that are complementary to the target, although in certain cases modified oligonucleotides as short as 7 bases in length have been used successfully as antisense oligonucleotides (Wagner et al., 1996). Most preferably, the antisense oligonucleotides
25 comprise a complementary sequence of 20-30 bases. Although oligonucleotides may be chosen that are antisense to any region of the gene or mRNA transcripts, in preferred embodiments the antisense oligonucleotides correspond to N-terminal or 5' upstream sites such as translation initiation, transcription initiation or promoter sites. In addition, 3'-untranslated regions may be targeted. Targeting to mRNA splicing sites has also been used
30 in the art but may be less preferred if alternative mRNA splicing occurs. In addition, the antisense is targeted, preferably, to sites in which mRNA secondary structure is not expected (see, e.g., Sainio et al., 1994) and at which proteins are not expected to bind. Finally, although the listed sequences are cDNA sequences, one of ordinary skill in the art may easily

derive the genomic DNA corresponding to the cDNA of a breast cancer expressed polypeptide. Thus, the present invention also provides for antisense oligonucleotides which are complementary to the genomic DNA corresponding to breast cancer expressed nucleic acids. Similarly, the use of antisense to allelic or homologous cDNAs and genomic DNAs are enabled without undue experimentation.

In one set of embodiments, the antisense oligonucleotides of the invention may be composed of "natural" deoxyribonucleotides, ribonucleotides, or any combination thereof. That is, the 5' end of one native nucleotide and the 3' end of another native nucleotide may be covalently linked, as in natural systems, via a phosphodiester internucleoside linkage. These oligonucleotides may be prepared by art-recognized methods, which may be carried out manually or by an automated synthesizer. They also may be produced recombinantly by vectors.

In preferred embodiments, however, the antisense oligonucleotides of the invention also may include "modified" oligonucleotides. That is, the oligonucleotides may be modified in a number of ways which do not prevent them from hybridizing to their target but which enhance their stability or targeting or which otherwise enhance their therapeutic effectiveness. The term "modified oligonucleotide" as used herein describes an oligonucleotide in which (1) at least two of its nucleotides are covalently linked via a synthetic internucleoside linkage (i.e., a linkage other than a phosphodiester linkage between the 5' end of one nucleotide and the 3' end of another nucleotide) and/or (2) a chemical group not normally associated with nucleic acids has been covalently attached to the oligonucleotide. Preferred synthetic internucleoside linkages are phosphorothioates, alkylphosphonates, phosphorodithioates, phosphate esters, alkylphosphonothioates, phosphoramidates, carbamates, carbonates, phosphate triesters, acetamides, carboxymethyl esters, and peptides.

The term "modified oligonucleotide" also encompasses oligonucleotides with a covalently modified base and/or sugar. For example, modified oligonucleotides include oligonucleotides having backbone sugars that are covalently attached to low molecular weight organic groups other than a hydroxyl group at the 3' position and other than a phosphate group at the 5' position. Thus modified oligonucleotides may include a 2'-O-alkylated ribose group. In addition, modified oligonucleotides may include sugars such as arabinose instead of ribose. The present invention, thus, contemplates pharmaceutical preparations containing modified antisense molecules that are complementary to and

hybridizable with, under physiological conditions, breast cancer expressed nucleic acids, together with pharmaceutically acceptable carriers.

Antisense oligonucleotides may be administered as part of a pharmaceutical composition. Such a pharmaceutical composition may include the antisense oligonucleotides in combination with any standard physiologically and/or pharmaceutically acceptable carriers which are known in the art. The compositions should be sterile and contain a therapeutically effective amount of the antisense oligonucleotides in a unit of weight or volume suitable for administration to a patient. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredients. The term "physiologically acceptable" refers to a non-toxic material that is compatible with a biological system such as a cell, cell culture, tissue, or organism. The characteristics of the carrier will depend on the route of administration. Physiologically and pharmaceutically acceptable carriers include diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials, which are well known in the art.

Expression of breast cancer nucleic acid molecules can also be determined using protein measurement methods to determine expression of SEQ ID NOs:1-51, e.g., by determining the expression of polypeptides encoded by SEQ ID NOs:1-51 (SEQ ID NOs: 52-102, respectively). Preferred methods of specifically and quantitatively measuring proteins include, but are not limited to: mass spectroscopy-based methods such as surface enhanced laser desorption ionization (SELDI; e.g., Ciphergen ProteinChip System), non-mass spectroscopy-based methods, antibody-capture protein arrays and immunohistochemistry-based methods such as 2-dimensional gel electrophoresis.

SELDI methodology may be used, through procedures known to those of ordinary skill in the art, to vaporize microscopic amounts of tumor protein and to create a "fingerprint" of individual proteins, thereby allowing simultaneous measurement of the abundance of many proteins in a single sample. Preferably SELDI-based assays may be utilized to classify breast cancer tumors. Such assays preferably include, but are not limited to the following examples. Gene products discovered by RNA microarrays may be selectively measured by specific (antibody mediated) capture to the SELDI protein disc (e.g., selective SELDI). Gene products discovered by protein screening (e.g., with 2-D gels), may be resolved by "total protein SELDI" optimized to visualize those particular markers of interest from among SEQ ID NOs:1-51. Predictive models of tumor classification from SELDI measurement of multiple markers from among SEQ ID NOs:1-51 may be utilized for the SELDI strategies. In an

additional embodiment a set of primary lymph node-negative premenopausal breast cancer tissues may be preferably utilized to determine the risk classification of breast cancer based on SELDI results.

The invention also involves agents such as polypeptides that bind to breast cancer-associated polypeptides, i.e., SEQ ID NOs:52-102. Such binding agents can be used, for example, in screening assays to detect the presence or absence of breast cancer-associated polypeptides and complexes of breast cancer-associated polypeptides and their binding partners and in purification protocols to isolate breast cancer-associated polypeptides and complexes of breast cancer-associated polypeptides and their binding partners. Such agents also may be used to inhibit the native activity of the breast cancer-associated polypeptides, for example, by binding to such polypeptides.

The invention, therefore, embraces peptide binding agents which, for example, can be antibodies or fragments of antibodies having the ability to selectively bind to breast cancer-associated polypeptides. Antibodies include polyclonal and monoclonal antibodies, prepared according to conventional methodology.

Significantly, as is well-known in the art, only a small portion of an antibody molecule, the paratope, is involved in the binding of the antibody to its epitope (see, in general, Clark, W.R. (1986) The Experimental Foundations of Modern Immunology Wiley & Sons, Inc., New York; Roitt, I. (1991) Essential Immunology, 7th Ed., Blackwell Scientific Publications, Oxford). The pFc' and Fc regions, for example, are effectors of the complement cascade but are not involved in antigen binding. An antibody from which the pFc' region has been enzymatically cleaved, or which has been produced without the pFc' region, designated an F(ab')₂ fragment, retains both of the antigen binding sites of an intact antibody. Similarly, an antibody from which the Fc region has been enzymatically cleaved, or which has been produced without the Fc region, designated an Fab fragment, retains one of the antigen binding sites of an intact antibody molecule. Proceeding further, Fab fragments consist of a covalently bound antibody light chain and a portion of the antibody heavy chain denoted Fd. The Fd fragments are the major determinant of antibody specificity (a single Fd fragment may be associated with up to ten different light chains without altering antibody specificity) and Fd fragments retain epitope-binding ability in isolation.

Within the antigen-binding portion of an antibody, as is well-known in the art, there are complementarity determining regions (CDRs), which directly interact with the epitope of the antigen, and framework regions (FRs), which maintain the tertiary structure of the

paratope (see, in general, Clark, 1986; Roitt, 1991). In both the heavy chain Fd fragment and the light chain of IgG immunoglobulins, there are four framework regions (FR1 through FR4) separated respectively by three complementarity determining regions (CDR1 through CDR3). The CDRs, and in particular the CDR3 regions, and more particularly the heavy chain CDR3, are largely responsible for antibody specificity.

It is now well-established in the art that the non-CDR regions of a mammalian antibody may be replaced with similar regions of conspecific or heterospecific antibodies while retaining the epitopic specificity of the original antibody. This is most clearly manifested in the development and use of "humanized" antibodies in which non-human CDRs are covalently joined to human FR and/or Fc/pFc' regions to produce a functional antibody. See, e.g., U.S. patents 4,816,567, 5,225,539, 5,585,089, 5,693,762 and 5,859,205.

Fully human monoclonal antibodies also can be prepared by immunizing mice transgenic for large portions of human immunoglobulin heavy and light chain loci. Following immunization of these mice (e.g., XenoMouse (Abgenix), HuMAb mice (Medarex/GenPharm)), monoclonal antibodies can be prepared according to standard hybridoma technology. These monoclonal antibodies will have human immunoglobulin amino acid sequences and therefore will not provoke human anti-mouse antibody (HAMA) responses when administered to humans.

Thus, as will be apparent to one of ordinary skill in the art, the present invention also provides for F(ab')₂, Fab, Fv and Fd fragments; chimeric antibodies in which the Fc and/or FR and/or CDR1 and/or CDR2 and/or light chain CDR3 regions have been replaced by homologous human or non-human sequences; chimeric F(ab')₂ fragment antibodies in which the FR and/or CDR1 and/or CDR2 and/or light chain CDR3 regions have been replaced by homologous human or non-human sequences; chimeric Fab fragment antibodies in which the FR and/or CDR1 and/or CDR2 and/or light chain CDR3 regions have been replaced by homologous human or non-human sequences; and chimeric Fd fragment antibodies in which the FR and/or CDR1 and/or CDR2 regions have been replaced by homologous human or non-human sequences. The present invention also includes so-called single chain antibodies.

Thus, the invention involves polypeptides of numerous size and type that bind specifically to polypeptides selected from SEQ ID NOs:52-102, and complexes of both breast cancer-associated polypeptides and their binding partners. These polypeptides may be derived also from sources other than antibody technology. For example, such polypeptide binding agents can be provided by degenerate peptide libraries which can be readily prepared

in solution, in immobilized form or as phage display libraries. Combinatorial libraries also can be synthesized of peptides containing one or more amino acids. Libraries further can be synthesized of peptoids and non-peptide synthetic moieties.

Phage display can be particularly effective in identifying binding peptides useful according to the invention. Briefly, one prepares a phage library (using e.g. m13, fd, or lambda phage), displaying inserts from 4 to about 80 amino acid residues using conventional procedures. The inserts may represent, for example, a completely degenerate or biased array. One then can select phage-bearing inserts which bind to the breast cancer-associated polypeptide. This process can be repeated through several cycles of reselection of phage that bind to the breast cancer-associated polypeptide. Repeated rounds lead to enrichment of phage bearing particular sequences. DNA sequence analysis can be conducted to identify the sequences of the expressed polypeptides. The minimal linear portion of the sequence that binds to the breast cancer-associated polypeptide can be determined. One can repeat the procedure using a biased library containing inserts containing part or all of the minimal linear portion plus one or more additional degenerate residues upstream or downstream thereof. Yeast two-hybrid screening methods also may be used to identify polypeptides that bind to the breast cancer-associated polypeptides.

Thus, the breast cancer-associated polypeptides of the invention, including fragments thereof, can be used to screen peptide libraries, including phage display libraries, to identify and select peptide binding partners of the breast cancer-associated polypeptides of the invention. Such molecules can be used, as described, for screening assays, for purification protocols, for interfering directly with the functioning of breast cancer-associated polypeptides and for other purposes that will be apparent to those of ordinary skill in the art. For example, isolated breast cancer-associated polypeptides can be attached to a substrate (e.g., chromatographic media, such as polystyrene beads, a filter, or an array substrate), and then a solution suspected of containing the binding partner may be applied to the substrate. If a binding partner that can interact with breast cancer-associated polypeptides is present in the solution, then it will bind to the substrate-bound breast cancer-associated polypeptide. The binding partner then may be isolated.

As detailed herein, the foregoing antibodies and other binding molecules may be used for example, to identify tissues expressing protein or to purify protein. Antibodies also may be coupled to specific diagnostic labeling agents for imaging of cells and tissues that express breast cancer-associated polypeptides or to therapeutically useful agents according to

standard coupling procedures. Diagnostic agents include, but are not limited to, barium sulfate, iocetamic acid, iopanoic acid, ipodate calcium, diatrizoate sodium, diatrizoate meglumine, metrizamide, tyropanoate sodium and radiodiagnostics including positron emitters such as fluorine-18 and carbon-11, gamma emitters such as iodine-123,

5 technitium-99m, iodine-131 and indium-111, nuclides for nuclear magnetic resonance such as fluorine and gadolinium.

The invention further includes protein microarrays for analyzing expression of breast cancer-associated peptides selected from SEQ ID NOs:52-102. In this aspect of the invention, standard techniques of microarray technology are utilized to assess expression of
10 the breast cancer-associated polypeptides and/or identify biological constituents that bind such polypeptides. The constituents of biological samples include antibodies, lymphocytes (particularly T lymphocytes), and the like. Protein microarray technology, which is also known by other names including: protein chip technology and solid-phase protein array technology, is well known to those of ordinary skill in the art and is based on, but not limited
15 to, obtaining an array of identified peptides or proteins on a fixed substrate, binding target molecules or biological constituents to the peptides, and evaluating such binding. See, e.g., G. MacBeath and S.L. Schreiber, "Printing Proteins as Microarrays for High-Throughput Function Determination," *Science* 289(5485):1760-1763, 2000.

Preferably antibodies or antigen binding fragments thereof that specifically bind
20 polypeptides selected from the group consisting of SEQ ID NOs:52-102 are attached to the microarray substrate in accordance with standard attachment methods known in the art. These arrays can be used to quantify the expression of the polypeptides identified herein.

In some embodiments of the invention, one or more control peptide or protein molecules are attached to the substrate. Preferably, control peptide or protein molecules
25 allow determination of factors such as peptide or protein quality and binding characteristics, reagent quality and effectiveness, hybridization success, and analysis thresholds and success.

The use of such methods to determine expression of breast cancer nucleic acids from among SEQ ID NOs:1-51 and/or proteins from among SEQ ID Nos:52-102 can be done with routine methods known to those of ordinary skill in the art and the expression determined by
30 protein measurement methods may be correlated to MAI levels and used as a prognostic method for selecting treatment strategies for breast cancer patients.

Examples

Introduction

To establish a prognostic tool for designing breast cancer treatment regimens,
5 expression patterns in primary breast cancer specimens were assessed and correlated with
clinical outcome. Primary breast cancer tumors from premenopausal women with no lymph
node metastases at the time of initial presentation were classified using the Mitotic Activity
Index (MAI), which has been shown to predict disease-free survival in this type of disease.
RNA was isolated, hybridized with Affymetrix HuFL human expression arrays, and analyzed
10 to ascertain which genes discriminate the two groups.

Methods

Breast Cancers Used for RNA Microarray Expression Analysis

Primary frozen breast cancers from premenopausal women with no lymph node
15 metastases at the time of initial presentation were assembled from material discarded
following routine surgical removal for diagnostic purposes. Institutional review and human
subjects approval for this project was obtained from Brigham and Women's Hospital. Fresh
tissue was frozen in liquid nitrogen, and a single fragment split for confirmatory histology
and RNA isolation. Individual fragments of frozen tumor tissues (estimated as 500 mg
20 minimum) were split by fracturing under liquid nitrogen, and a portion processed for
confirmatory histology using standard methods. The remaining tissue was used for
synchronous RNA, protein, and DNA isolations with TRIzol reagents (Life Technologies,
Inc., Rockville, MD) using standard methods. Only tumors where the actual frozen tissue
contained >50% tumor cells were used.

Mitotic Activity Index

All tumors were classified by Mitotic Activity Index (Baak et al., 1989; van Diest et
al., 1991; van Diest et al., 1992(a); Uytendinck et al., 1990; van Diest et al., 1992(b); Jannink
et al., 1996; Baak et al., 1992; Baak et al., 1993) using paraffin H&E stained tissues sections
30 prepared for diagnostic purposes at the time of excision. The MAI is the total number of
mitoses counted in 10 consecutive high-power fields (objective, x40; numeric aperture, 0.75;
field diameter, 450 microns) in the most cellular area at the periphery of the tumor, with the
subjectively highest mitotic activity (Jannink et al., 1995). Risk groups have previously been

defined using a threshold of 10 mitoses/unit area (Tosi et al., 1986; Jannink et al., 1995; Theissig et al., 1996). Tumors with $MAI \geq 10$ were assigned to the high risk group, and those with $MAI \leq 3$ to the low risk group.

5 *Microarray Expression Analysis*

RNA from 27 qualifying tumors was reverse transcribed and resultant cDNA used for *in vitro* transcriptional synthesis of fluorescently labeled nucleic acid probes which were then hybridized to Affymetrix HuFL human expression arrays (approximately 7100, probe sets, estimated 5800 unique genes). Hybridization images were analyzed with Affymetrix
10 software to generate a data matrix of named probes by quantitative expression level in each tissue. RNA labeling, microarray hybridization, and microarray analysis were performed as per vendor's instructions for HuGeneFL array (Affymetrix, Santa Clara, CA). Four tumors were excluded from analysis because they failed to meet quality control criteria for microarray hybridization: 3 cases had low hybridization signal, one case had high
15 background.

Results

Analysis of 23 primary breast cancer specimens from premenopausal lymph node negative women were split between two prognostic groups (Low MAI, $MAI \leq 3$, $n=11$ and
20 High MAI, $MAI \geq 10$, $n=12$) and was accomplished as follows. Affymetrix HuFL expression values were normalized by scaling so the sum of AD (AD units are the quantitative expression units used by Affymetrix) values in each sample was 3,000,000; genes for which RNA abundance was absent or marginal were reset to a value of 0, then any values less than 20 were reset to 20. The result is the GPT datastate, which was then log transformed and
25 discriminating genes selected by t-test comparison of the logged data between low and high MAI groups. Significance cutoffs for the t-tests used Permax < 0.96 based on 10,000 random permutations of the data. Permax is a data analysis software tool for testing the significance of gene expression. It has been presented by Mutter, et al., 8th International Workshop on Chromosomes in Solid Tumors, Tucson, AZ, 2000; and is available online at
30 biowww.dfci.harvard.edu/~gray/permax.html and from Robert J. Gray, Department of Biostatistical Science, Dana-Farber Cancer Institute, 44 Binney Street Boston, MA 02115. Permax details enclosed therein are incorporated by reference herein. Seventy eight of 7070 Affymetrix probe sets were selected by Permax.

Filters for minimum divergence between the average expression values of the two groups (Low vs. High MAI) were applied as follows: ratio of means ≥ 2 , and difference between means ≥ 100 . It was determined that 51/78 genes passed these filters. The final 51 selected genes which discriminate between low and high MAI subgroups appear in Table 1 and as SEQ ID NOs:1-51. Average expression in high MAI tumors and low MAI tumors is shown as HX and LX, respectively.

Table 1. Gene list identifying 51 genes that discriminate low from high MAI breast cancers.

SEQ ID NO	Short Name	GenBank Acc.No.	Permax	HX	LX	FOLDABS	DIFFABS
1	ABCB2	X57522	0.9577	501	83	6.0	417
2	ACTA2	X13839	0.7131	3098	6152	2.0	3054
3	AMD1	M21154	0.0808	257	50	5.1	207
4	APM2	D45370	0.3317	590	2682	4.5	2092
5	ASAH	U70063	0.8435	360	990	2.8	630
6	BARD1	U76638	0.5637	242	102	2.4	140
7	CCNH	U11791	0.9104	104	204	2.0	100
8	CCT2	U91327	0.8801	280	109	2.6	171
9	CDC20	U05340	0.0669	579	20	29.0	559
10	CDC34	L22005	0.6979	182	41	4.4	141
11	CDKN3	U02681	0.0072	454	63	7.2	391
12	CKS1	X54941	0.8823	539	219	2.5	320
13	CKS2	X54942	0.1881	413	119	3.5	294
14	COX7A1	M83186	0.9223	89	326	3.6	236
15	CPA3	M73720	0.8234	132	357	2.7	225
16	CPE	X51405	0.1984	80	243	3.0	163
17	CX3CR1	U20350	0.0317	70	328	4.7	258
18	DLG4	U83192	0.3427	20	179	8.9	159
19	DOC1	U53445	0.927	122	276	2.3	154
20	DXS9879E	X92896	0.9448	744	331	2.3	413
21	E2-EPF	M91670	0.9602	324	20	16.2	304
22	ElastinAlt2	U77846	0.8368	417	2210	5.3	1792
23	GTF2A1	U14193	0.7495	528	249	2.1	279
24	GUA5MPST	U10860	0.6129	599	114	5.2	485
25	H2AFX	X14850	0.8106	496	193	2.6	303
26	H2BFA	M60750	0.2334	508	143	3.6	365
27	Hevin	X86693	0.7484	529	1686	3.2	1157
28	HNRPH2	U01923	0.9056	106	231	2.2	126
29	HPV16E1Bind	U96131	0.2439	194	78	2.5	116
30	IDUA	M74715	0.1712	176	594	3.4	418
31	IGF1	X57025	0.9213	79	265	3.4	186
32	IQGAP2	U51903	0.9517	137	321	2.3	184
33	ISG15	M13755	0.9316	2133	386	5.5	1747
34	JAG1	U61276	0.9466	79	264	3.3	185
35	LAMA2	Z26653	0.8882	31	213	6.8	182
36	LAMB2	X79683	0.083	156	658	4.2	502
37	LBR	L25931	0.5991	221	68	3.2	153
38	MMP2	M55593	0.93	1765	3670	2.1	1905
39	MMSDH	M93405	0.9072	297	669	2.3	372
40	MYH11	AF001548	0.3109	164	777	4.7	612
41	MYLK	U48959	0.8351	158	680	4.3	522
42	PDE4A	L20965	0.8912	34	176	5.2	142
43	SCNN1A	X76180	0.694	352	864	2.5	511
44	SCYB10	X02530	0.4416	528	83	6.4	445
45	SNRPB	X17567	0.8965	1473	638	2.3	835
46	STAT1	M97936	0.9553	440	20	22.0	420
47	TAF2A	X07024	0.6819	193	65	2.9	127
48	TCEAL1	M99701	0.5595	241	749	3.1	508
49	TPM1	Z24727	0.5676	1266	2533	2.0	1267
50	TPS2	M33493	0.3638	194	892	4.6	698
51	UBCH10	U73379	0.1972	1519	639	2.4	880

Several features of selected genes provide reassurance that low frequency random events were not the cause of expression differences between groups. A review of the 51 selected genes (Table 1) shows that five pairs of genes known to be co-expressed were selected independently (two carboxypeptidases, two histones, two cdc28, two ubiquitins, two laminins, and myosin/tropomyosin), and reciprocal regulation of ligand and receptor, a common regulatory pattern, occurred once (laminin and lamin receptor) amongst genes selected.

The first expectation is that genes whose expression is linked to cell division would be represented in this comparison of tumors whose mitotic activity differs systematically. This was in fact the largest category of selected genes, with expression of 11/12 cell cycle genes greatest in the high MAI group. Genes which are preferentially expressed (at higher levels) in the low MAI group include those encoding extracellular matrix or enzymes which may remodel extracellular matrix (proteolytic enzymes).

The gene expression data presented in Table 1 can be used to generate an expression matrix of 51 selected genes by 23 tissues examined. Using standard clustering algorithms, dendrograms can be provided on the borders of the matrix (e.g., using Wards linkage and Euclidean distance) to show cluster relationships between tissues and genes. Similarly, a gene expression matrix can be generated using data normalized by standard deviation for each gene [STD(GPT)]. Dendrograms on borders of the matrix can be provided to show cluster relationships between tissues and genes. In this type of matrix, clustering of genes is based upon relative changes without bias due to absolute expression level, because each gene is expressed in standard deviation from the mean for that specific gene. However, unlike the other expression matrix described above, the absolute magnitude of expression cannot be directly inferred from this plot.

References

Harrison's Principles of Internal Medicine, 14/e, (1998) McGraw-Hill Companies, New York.

Jannink, I, van Diest, P.J., Baak, J.P. (1995) Comparison of the prognostic value of four methods to assess mitotic activity in 186 invasive breast cancer patients: classical and random

mitotic activity assessments with correction for volume percentage of epithelium. Hum Pathol Oct;26(10):1086-92.

5 Baak JP, van Diest PJ, Ariens AT, van Beek MW, Bellot SM, Fijnheer J, van Gorp LH, Kwee WS, Los J, Peterse HC: The Multicenter Morphometric Mammary Carcinoma Project (MMMCP). A nationwide prospective study on reproducibility and prognostic power of routine quantitative assessments in The Netherlands. Pathol Res Pract 1989, 185:664-670.

10 van Diest PJ, Baak JP: The morphometric prognostic index is the strongest prognosticator in premenopausal lymph node-negative and lymph node-positive breast cancer patients. Hum Pathol 1991, 22:326-330.

15 van Diest PJ, Baak JP, Matze-Cok P, Bacus SS: Prediction of response to adjuvant chemotherapy in premenopausal lymph node positive breast cancer patients with morphometry, DNA flow cytometry and HER-2/neu oncoprotein expression. Preliminary results. Pathol Res Pract 1992, 188:344-349.

20 Uyterlinde AM, Baak JP, Schipper NW, Peterse H, Matze E, Meijer CJ: Further evaluation of the prognostic value of morphometric and flow cytometric parameters in breast-cancer patients with long follow-up. Int J Cancer 1990, 45:1-7.

25 van Diest PJ, Baak JP, Matze-Cok P, Wisse-Brekelmans EC, van Galen CM, Kurver PH, Bellot SM, Fijnheer J, van Gorp LH, Kwee WS: Reproducibility of mitosis counting in 2,469 breast cancer specimens: results from the Multicenter Morphometric Mammary Carcinoma Project. Hum Pathol 1992, 23:603-607.

Jannink I, Risberg B, van Diest PJ, Baak JP: Heterogeneity of mitotic activity in breast cancer. Histopathology 1996, 29:421-428.

30 Baak JP, Wisse-Brekelmans EC, Kurver PH, van Gorp LH, Voorhorst FJ, Miettinen OS: Regional differences in breast cancer survival are correlated with differences in differentiation and rate of proliferation. Hum Pathol 1992, 23:989-992.

Baak JP, van Diest PJ, Benraad T, Matze-Cok E, Brugghe J, Schuurmans LT, Littooy JJ: The Multi-Center Morphometric Mammary Carcinoma Project (MMMCP) in The Netherlands: value of morphometrically assessed proliferation and differentiation. J Cell Biochem Suppl 1993, 17G:220-225.

5

Tosi P, Luzi P, Sforza V, Santopietro R, Bindi M, Tucci E, Barbini P, Baak JP: Correlation between morphometrical parameters and disease-free survival in ductal breast cancer treated only by surgery. Appl Pathol 1986, 4:33-42.

10

Theissig F, Baak JP, Schuurmans L, Haroske G, Meyer W, Kunze KD: 'Blind' multicenter evaluation of the prognostic value of DNA image cytometric and morphometric features in invasive breast cancer. Anal Cell Pathol 1996, 10:85-99.

The Chipping Forecast (1999) Nature Genetics, 21(1):1-60.

15

Gwynne, P., and Page, G., (1999) Microarray Analysis: the next revolution in Molecular Biology, Science eMarketplace, Science, August 6. (sciencemag.org/feature/e-market/benchtop/micro.shl)

20

Wagner et al., *Nature Biotechnol.* (1996) 14:840-844.

Sainio, K., Saarma, M., Nonclercq, D., Paulin, L., and Sariola, H. (1994) Antisense inhibition of low-affinity nerve growth factor receptor in kidney cultures: power and pitfalls. Cell Mol. Neurobiol. 14(5):439-457.

25

Molecular Cloning: A Laboratory Manual, (1989) J. Sambrook, et al., eds., Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Current Protocols in Molecular Biology, (1999) F.M. Ausubel, et al., eds., John Wiley & Sons, Inc., New York.

30

Mutter, G.L., Baak, J.P.A., Cai, T., Fitzgerald, J., Gray, R., Gentleman, R., Gullans, S., Ibrahim, J., Neuberg, D., and Wilcox, M. Altered Gene Expression in Endometrioid

Endometrial Adenocarcinomas Analyzed by High Density Microarrays. 8th International Workshop on Chromosomes in Solid Tumors (Tucson,AZ) . 2000.

The present invention is not limited in scope by the examples provided, since the
5 examples are intended as illustrations of various aspects of the invention and other
functionally equivalent embodiments are within the scope of the invention. Various
modifications of the invention in addition to those shown are described herein will become
apparent to those skilled in the art for the foregoing description and fall within the scope of
the appended claims. The advantages and objects of the invention are not necessarily
10 encompassed by each embodiment of the invention. All references, patents, and patent
publications that are recited in this application are incorporated in their entirety herein by
reference.

We claim:

Claims

1. A method for diagnosing breast cancer in a subject suspected of having breast cancer comprising:
 - 5 obtaining from the subject a breast tissue sample suspected of being cancerous,
determining the expression of a set of nucleic acid molecules or expression products thereof in the breast tissue sample, wherein the set of nucleic acid molecules comprises at least two nucleic acid molecules selected from the group consisting of SEQ ID NOs:1-51.
- 10 2. The method of claim 1, wherein the set of nucleic acid molecules comprises at least 3 nucleic acid molecules selected from the group consisting of SEQ ID NOs:1-51.
3. The method of claim 1, wherein the set includes at least 4 nucleic acid molecules selected from the group consisting of SEQ ID NOs:1-51.
- 15 4. The method of claim 1, wherein the set includes at least 5 nucleic acid molecules selected from the group consisting of SEQ ID NOs:1-51.
5. The method of claim 1, wherein the set includes at least 10 nucleic acid molecules selected from the group consisting of SEQ ID NOs:1-51.
- 20 6. The method of claim 1, wherein the set includes at least 15 nucleic acid molecules selected from the group consisting of SEQ ID NOs:1-51.
7. The method of claim 1, wherein the set includes at least 20 nucleic acid molecules selected from the group consisting of SEQ ID NOs:1-51.
8. The method of claim 1, wherein the set includes at least 30 nucleic acid molecules selected from the group consisting of SEQ ID NOs:1-51.
- 30 9. The method of claim 1, wherein the set includes at least 40 nucleic acid molecules selected from the group consisting of SEQ ID NOs:1-51.

10. The method of claim 1, further comprising:

determining the expression of the set of nucleic acid molecules or expression products thereof in a non-cancerous breast tissue sample, and comparing the expression of the set of nucleic acid molecules or expression products thereof in the breast tissue sample suspected of
5 being cancerous and the non-cancerous breast tissue sample.

11. A method for identifying a set of nucleic acid markers or expression products thereof effective for determining the prognosis of cancer, comprising:

obtaining a plurality of tumor tissue samples from a plurality of subjects afflicted with
10 cancer,

classifying the plurality of tumor tissue samples according to mitotic activity index (MAI) into high MAI and low MAI groups,

determining differences in the expression of a plurality of nucleic acid molecules or expression products thereof in the tumor tissue samples, and

15 selecting as a set of nucleic acid markers the nucleic acid molecules or expression products thereof which are differentially expressed in the high MAI and the low MAI groups,

wherein the set of nucleic acid markers or expression products thereof effective for determining poor prognosis of cancer comprises one or more nucleic acid molecules or expression products thereof which are preferentially expressed in high MAI tumor tissue
20 samples, and wherein the set of nucleic acid markers or expression products thereof effective for determining good prognosis of cancer comprises one or more nucleic acid molecules or expression products thereof which are preferentially expressed in low MAI tumor tissue samples.

25 12. The method of claim 11, wherein the cancer is breast cancer.

13. The method of claim 11, wherein the differences in the expression of a plurality of nucleic acid molecules are determined by a method selected from the group consisting of nucleic acid hybridization and nucleic acid amplification.

30 14. The method of claim 13, wherein the nucleic acid hybridization is performed using a solid-phase nucleic acid molecule array.

-36-

15. A method for selecting a course of treatment of a subject having or suspected of having cancer, comprising:

obtaining from the subject a tissue sample suspected of being cancerous,

determining the expression of a set of nucleic acid markers or expression products thereof which are differentially expressed in high MAI tumor tissue samples to determine the MAI of the tissue sample of the subject, and

selecting a course of treatment appropriate to the cancer of the subject.

16. The method of claim 15 wherein the cancer is breast cancer.

17. The method of claim 16, further comprising:

determining the expression of a set of nucleic acid markers that are differentially expressed in low MAI breast tumor tissue samples.

18. The method of claim 15, wherein the expression of a set of nucleic acid markers is determined by a method selected from the group consisting of nucleic acid hybridization and nucleic acid amplification.

19. The method of claim 18, wherein the nucleic acid hybridization is performed using a solid-phase nucleic acid molecule array.

20. A method for evaluating treatment of cancer, comprising:

obtaining a first determination of the expression of a set of nucleic acid molecules or expression products thereof, which are differentially expressed in high MAI tumor tissue samples to determine the MAI of the tissue sample from a subject undergoing treatment for cancer,

obtaining a second determination of the expression of a set of nucleic acid molecules or expression products thereof, which are differentially expressed in high MAI tumor tissue samples to determine the MAI of the second tissue sample from the subject after obtaining the first determination,

comparing the first determination of expression to the second determination of expression as an indication of evaluation of the treatment.

21. The method of claim 20, wherein the cancer is breast cancer.

22. The method of claim 21, further comprising:

determining the expression of a set of nucleic acid markers which are differentially
5 expressed in low MAI breast tumor tissue samples.

23. The method of claim 20, wherein the expression of a set of nucleic acid markers is
determined by a method selected from the group consisting of nucleic acid hybridization and
nucleic acid amplification.

10 24. The method of claim 20, wherein the nucleic acid hybridization is performed using a
solid-phase nucleic acid molecule array.

25. A solid-phase nucleic acid molecule array consisting essentially of at least two nucleic
15 acid molecules selected from the group consisting of SEQ ID NOs:1-51 fixed to a solid
substrate.

26. The solid-phase nucleic acid molecule array of claim 24, further comprising at least
one control nucleic acid molecule.

20 27. The solid-phase nucleic acid molecule array of claim 24, wherein the set of nucleic
acid molecules comprises at least 3 nucleic acid molecules selected from the group consisting
of SEQ ID NOs:1-51.

25 28. The solid-phase nucleic acid molecule array of claim 24, wherein the set includes at
least 4 nucleic acid molecules selected from the group consisting of SEQ ID NOs:1-51.

29. The solid-phase nucleic acid molecule array of claim 24, wherein the set includes at
least 5 nucleic acid molecules selected from the group consisting of SEQ ID NOs:1-51.

30 30. The solid-phase nucleic acid molecule array of claim 24, wherein the set includes at
least 10 nucleic acid molecules selected from the group consisting of SEQ ID NOs:1-51.

31. The solid-phase nucleic acid molecule array of claim 24, wherein the set includes at least 15 nucleic acid molecules selected from the group consisting of SEQ ID NOs:1-51.

32. The solid-phase nucleic acid molecule array of claim 24, wherein the set includes at least 20 nucleic acid molecules selected from the group consisting of SEQ ID NOs:1-51.

33. The solid-phase nucleic acid molecule array of claim 24, wherein the set includes at least 30 nucleic acid molecules selected from the group consisting of SEQ ID NOs:1-51.

34. The solid-phase nucleic acid molecule array of claim 24, wherein the set includes at least 40 nucleic acid molecules selected from the group consisting of SEQ ID NOs:1-51.

35. The solid-phase nucleic acid molecule array of claim 24, wherein the solid substrate comprises a material selected from the group consisting of glass, silica, aluminosilicates, borosilicates, metal oxides such as alumina and nickel oxide, various clays, nitrocellulose, and nylon.

36. The solid-phase nucleic acid molecule array of claim 24, wherein the nucleic acid molecules are fixed to the solid substrate by covalent bonding.

37. A solid-phase protein microarray comprising at least two antibodies or antigen-binding fragments thereof, that specifically bind at least two different polypeptides selected from the group consisting of SEQ ID NOs:52-102, fixed to a solid substrate.

38. The protein microarray of claim 37, wherein the microarray further comprises an antibody or antigen-binding fragment thereof, that binds specifically to a cancer-associated polypeptide other than those selected from the group consisting of SEQ ID NOs:52-102.

39. The protein microarray of claim 38, wherein the cancer-associated polypeptide other than those selected from the group consisting of SEQ ID NOs:52-102 is a breast cancer associated polypeptide.

40. The protein microarray of claim 37, further comprising at least one control polypeptide molecule.

41. The protein microarray of claim 37, wherein the antibodies are monoclonal or polyclonal antibodies.

42. The protein microarray of claim 37, wherein the antibodies are chimeric, human, or humanized antibodies.

43. The protein microarray of claim 37, wherein the antibodies are single chain antibodies.

44. The protein microarray of claim 37, wherein the antigen-binding fragments are F(ab')₂, Fab, Fd, or Fv fragments.

45. A method for identifying lead compounds for a pharmacological agent useful in the treatment of breast cancer, comprising:

contacting a breast cancer cell or tissue with a candidate pharmacological agent,

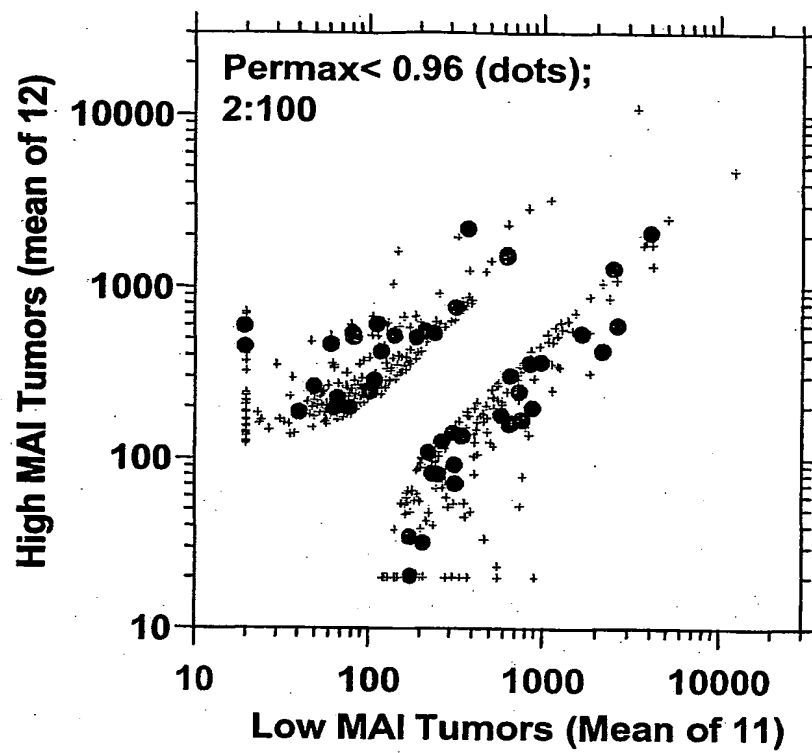
determining the expression of a set of nucleic acid molecules in the breast cancer cell

or tissue sample under conditions which, in the absence of the candidate pharmacological agent, permit a first amount of expression of the set of nucleic acid molecules wherein the set of nucleic acid molecules comprises at least two nucleic acid molecules selected from the group consisting of SEQ ID NOs:1-51, and

detecting a test amount of the expression of the set of nucleic acid molecules, wherein a decrease in the test amount of expression in the presence of the candidate pharmacological agent relative to the first amount of expression indicates that the candidate pharmacological agent is a lead compound for a pharmacological agent which is useful in the treatment of breast cancer.

46. The method of claim 45, wherein the set of nucleic acid molecules is differentially expressed in high MAI breast tumor tissue samples.

1/1

**Fig. 1**

- 1 -

SEQUENCE LISTING

<110> The Brigham and Women's Hospital, Inc
Baak, Jan

<120> Prognostic Classification of Breast Cancer

<130> B0801/7224WO(JRV)

<150> US 60/222,093

<151> 2000-07-28

<160> 102

<170> PatentIn version 3.0

<210> 1

<211> 2824

<212> DNA

<213> Homo Sapiens

<400> 1

```

gcggccgctt tcgatttcgc tttccctaa atggctgagc ttctcgccag cgcaggatca      60
gcctgttcct gggactttcc gagagccccg cctcgttcc ctccccagc cgccagtagg      120
ggaggactcg gcggtacccg gagcttcagg cccacccggg gcgaggagag tcccagacct      180
ggccgggacc gggacggcgt ccgagtgcc aatggctagct ctaggtgtcc cgtccccgc      240
gggtgcccgt gctcccccg agcttctctc gcatggctgg ggacagtact gctacttctc      300
gccgactggg tgctgctccg gaccgcgctg ccccgcatat tctccctgct ggtgcccacc      360
gcgctgccac tgctccgggt ctgggcgggt ggctgagcc gctgggcccgt gctctggctg      420
ggggcctgcg gggtcctcag ggcaacgggt ggctccaaga gcgaaaacgc aggtgccag      480
ggctggctgg ctgctttgaa gccattagct gcggcactgg gcttggccct gccgggactt      540
gccttgcttc gagagctgat ctcatgggga gccccgggt ccgaggatag caccaggcta      600
ctgcactggg gaagtcaccc taccgccttc gttgtcagtt atgcagcggc actgcccgca      660
gcagccctgt ggcacaaact cgggagcctc tgggtgcccg gcggtcaggg cggctctgga      720
aaccctgtgc gtcggcttct aggtgcctg ggctcggaga cgcgccgct ctcgctgttc      780
ctggtcctgg tggctctctc ctctcttggg gagatggcca ttccattctt tacgggcccgc      840
ctcactgact ggattctaca agatggctca gccgatacct tcaactgaaa cttactctc      900
atgtccattc tcaccatagc cagtgcagt ctggagtctg tgggtgacgg gatctataac      960
aacaccatgg gccacgtgca cagccacttg caggagaggg tgtttggggc tgcctgccc      1020
caggagacgg agtttttcca acagaaccag acaggtaaca tcatgtctcg ggtaacagag      1080
gacacgtcca ccctgagtga ttctctgagt gagaatctga gcttatttct gtggtacctg      1140

```

- 2 -

gtgcgaggcc tatgtctctt ggggatcatg ctctggggat cagtgtccct caccatggtc	1200
accctgatca ccctgcctct gcttttcctt ctgcccaaga aggtgggaaa atggtaccag	1260
ttgctggaag tgcaggtgcg ggaatctctg gcaaagtcca gccaggtggc cattgaggct	1320
ctgtcggccca tgcctacagt tcgaagcttt gccaacgagg agggcgaagc ccagaagttt	1380
agggaaaagc tgcaagaaat aaagacactc aaccagaagg aggctgtggc ctatgcagtc	1440
aactcctgga ccactagtat ttcaggatat ctgctgaaag tgggaatcct ctacattggt	1500
gggcagctgg tgaccagtgg ggctgtaagc agtgggaacc ttgtcacatt tgttctctac	1560
cagatgcagt tcacccaggc tgtggaggta ctgctctcca tctacccag agtacagaag	1620
gctgtgggct cctcagagaa aatatttgag tacctggacc gcacccctcg ctgccaccc	1680
agtggctctgt tgactccctt aacttggag ggcttgtcc agttccaaga tgtctccttt	1740
gcctacccaa accgcccaga tgtcttagtg ctacaggggc tgacattcac cctacgcct	1800
ggcgaggtga cggcgctggg gggacccaat gggctctggga agagcacagt ggctgccctg	1860
ctgcagaatc tgtaccagcc caccggggga cagctgctgt tggatgggaa gcccttccc	1920
caatatgagc accgctacct gcacaggcag gtggctgcag tgggacaaga gccacaggta	1980
tttggaaaga gtcttcaaga aaatattgcc tatggcctga cccagaagcc aactatggag	2040
gaaatcacag ctgctgcagt aaagtctggg gcccatagtt tcctctctgg actccctcag	2100
ggctatgaca cagaggtaga cgaggctggg agccagctgt caggggggtca gcgacaggca	2160
gtggcgttgg cccgagcatt gatccggaaa ccgtgtgtac ttatcctgga tgatgccacc	2220
agtgccttg atgcaaacag ccagttacag gtggagcagc tcctgtacga aagccctgag	2280
cggtactccc gctcagtget tctcatcacc cagcacctca gcctgggtgga gcaggctgac	2340
cacatcctct ttctggaagg aggcgtatc cgggaggggg gaaccaccca gcagctcatg	2400
gagaaaaagg ggtgctactg ggccatggtg caggctcctg cagatgctcc agaatgaaag	2460
ccttctcaga cctgcgcact ccctctccct cccttttctt ctctctgtgg tggagaacca	2520
cagctgcaga gtagcagctg cctccaggat gagttacttg aaatttgcct tgagtgtgtt	2580
acctccttcc caagctcctc gtgataatgc agacttctg gagtacaaac acaggatttg	2640
taattcctac tgtaacggag tttagagcca gggctgatgc tttggtgtgg ccagcactct	2700
gaaactgaga aatgttcaga atgtacggaa agatgatcag ctattttcaa cataactgaa	2760
ggcatatgct ggcccataaa caccctgtag gttcttgata tttataataa aattggtgtt	2820
ttgt	2824

- 3 -

<211> 1330

<212> DNA

<213> Homo Sapiens

<400> 2

```

gcagcccagc caagcactgt caggaatcct gtgaagcagc tccagctatg tgtgaagaag      60
aggacagcac tgccttggtg tgtgacaatg gctctgggct ctgtaaggcc ggctttgctg      120
gggacgatgc tcccagggtt gttttcccat ccattgtggg acgtcccaga catcaggggg      180
tgatggtggg aatgggacaa aaagacagct acgtgggtga cgaagcacag agcaaaagag      240
gaatcctgac cctgaagtac ccgatagaac atggcatcat caccaactgg gacgacatgg      300
aaaagatctg gcaccactct ttctacaatg agcttctgtg tgcccctgaa gagcatccca      360
ccctgctcac ggaggcaccc ctgaacccca aggccaaccg ggagaaaatg actcaaatta      420
tgtttgagac tttcaatgtc ccagccatgt atgtggctat ccaggcgggtg ctgtctctct      480
atgcctctgg acgcacaact ggcacgtgctc tggactctgg agatgggtgtc acccacaatg      540
tccccatcta tgagggtctat gccttgcccc atgccatcat gcgtctggat ctggctggcc      600
gagatctcac tgactacctc atgaagatcc tgactgagcg tggctattcc ttcgttacta      660
ctgctgagcg tgagattgtc cgggacatca aggagaaact gtgttatgta gctctggact      720
ttgaaaatga gatggccact gccgcacct catcctccct tgagaagagt tacgagttgc      780
ctgatgggca agtgatcacc atcggaatg aacgtttccg ctgcccagag accctgttcc      840
agccatcctt catcgggatg gagtctgctg gcattccatga aaccacctac aacagcatca      900
tgaagtgtga tattgacatc aggaaggacc tctatgctaa caatgtccta tcagggggca      960
ccactatgta ccctggcatt gccgaccgaa tgcagaagga gatcacggcc ctagcaccca     1020
gcaccatgaa gatcaagatc attgcccttc cggagcgcaa atactctgtc tggatcgggtg     1080
gtccatcctt ggctctctg tccaccttcc agcagatgtg gatcagcaaa caggaatacg     1140
atgaagccgg gccttcatt gtccaccgca aatgcttcta aaacactttc ctgctcctct     1200
ctgtctctag cacacaactg tgaatgtcct gtggaattat gccttcagtt cttttccaaa     1260
tcattcctag ccaaagctct gactcgttac ctatgtgttt ttttaataaat ctgaaatagg     1320
ctactggtaa                                     1330

```

<210> 3

<211> 1805

<212> DNA

<213> Homo Sapiens

<400> 3

```

aagagactga actgtatctg cctctatttc caaaagactc acgttcaact ttcgctcaca      60

```

- 4 -

caaagccggg aaaattttat tagtcctttt tttaaaaaaa gttaatatataa aattatagca	120
aaaaaaaaaa ggaacctgaa ctttagtaac acagctggaa caatcgcagc ggcggcggca	180
gcggcgggag aagaggttta atttagttga ttttctgtgg ttgttggttg ttcgctagtc	240
tcacggtgat ggaagctgca catttttttcg aagggaccga gaagctgctg gaggtttggt	300
tctcccggca gcagcccgaac gcaaaccaag gatctgggga tcttcgcact atcccaagat	360
ctgagtggga catacttttg aaggatgtgc aatgttcaat cataagtgtg aaaaaaactg	420
acaagcagga agcttatgta ctcaagtga gtagcatgtt tgtctccaag agacgtttca	480
ttttgaagac atgtggtacc accctcttgc tgaaagcact gggtcccctg ttgaagcttg	540
ctagggatta cagtgggttt gactcaattc aaagcttctt ttattctcgt aagaatttca	600
tgaagccttc tcaccaaggg taccacacacc ggaatttcca ggaagaaata gagtttctta	660
atgcaatttt cccaaatgga gcaggatatt gtatgggacg tatgaattct gactgttggt	720
acttatatac tctggatttc ccagagagtc gggtaatcag tcagccagat caaaccttgg	780
aaattctgat gagtgagctt gaccagcag ttatggacca gttctacatg aaagatgggtg	840
ttactgcaaa ggatgtcact cgtgagagtg gaattcgtga cctgatacca ggttctgtca	900
ttgatgccac aatgttcaat ccttgtgggt attcgatgaa tggaatgaaa tcggatggaa	960
cttattggac tattcacatc actccagaac cagaattttc ttatgttagc tttgaaacaa	1020
acttaagtca gacctcctat gatgacctga tcaggaaagt tgtagaagtc ttcaagccag	1080
gaaaatttgt gaccaccttg tttgttaatc agagttctaa atgtcgcaca gtgcttgctt	1140
cgccccagaa gattgaaggt ttttaagcgtc ttgattgcca gagtgctatg ttcaatgatt	1200
acaattttgt ttttaccagt tttgctaaga agcagcaaca acagcagagt tgattaagaa	1260
aatgaagaa aaaacgcaaa aagagaacac atgtagaagg tggatggatgc tttctagatg	1320
tcgatgctgg gggcagtgtc ttccataacc accactgtgt agttgcagaa agccctagat	1380
gtaatgatag tgtaatcatt ttgaattgta tgcattatta tatcaaggag ttagatatct	1440
tgcatgaatg ctctcttctg tgtttaggta ttctctgcca ctcttgctgt gaaattgaag	1500
tggatgtaga aaaaaccttt tactatatga aactttacaa cacttggtgaa agcaactcaa	1560
tttggtttat gcacagtgtg atatttctcc aagtatcatc caaaattccc cacagacaag	1620
gctttcgtcc tcattaggtg ttggcctcag cctaaccctc taggactgtt ctattaaatt	1680
gctgccagaa ttttaccatcc agttacctcc actttctaga acatattctt tactaatgtt	1740
attgaaacca atttctactt catactgatg tttttggaaa cagcaattaa agtttttctt	1800
ccatg	1805

- 5 -

<210> 4
 <211> 419
 <212> DNA
 <213> Homo Sapiens

<400> 4
 ctcttgacga ctccacagat accccaagc catggcaagc aagggcttgc aggacctgaa 60
 gcaacaggtg gaggggaccg cccaggaagc cgtgtcagcg gccggagcgg cagctcagca 120
 agtgggtggac caggccacag aggcggggca gaaagccatg gaccagctgg ccaagaccac 180
 ccaggaaacc atcgacaaga ctgctaacca ggcctctgac accttctctg ggatcgggaa 240
 aaaattcggc ctcttgaaat gacagcaggg agacttgggt cggcctcctg aaatgatagc 300
 agggagactt gggtgacccc cttccaggc gccatctagc acagcctggc cctgatctcc 360
 gggcagccac cacctcctcg gtctgcccc tcattaaaat tcacgttccc accctgaaa 419

<210> 5
 <211> 2333
 <212> DNA
 <213> Homo Sapiens

<400> 5
 ggcacgaggc tagagcgatg ccgggccgga gttgcgtcgc cttagtcctc ctggctgccg 60
 ccgtcagctg tgccgtcgcg cagcacgcgc cgccgtggac agaggactgc agaaaatcaa 120
 cctatcctcc ttcaggacca acgtacagag gtgcagttcc atggtacacc ataaatcttg 180
 acttaccacc ctacaaaaga tggcatgaat tgatgcttga caaggcacca atgctaaagg 240
 ttatagtga tttcttgaag aatatgataa atacattcgt gccagtggga aaagtatatgc 300
 aggtggtgga tgaaaaattg cctggcctac ttggcaactt tctggccct tttgaagagg 360
 aaatgaaggg tattgccgct gttactgata tacctttagg agagattatt tcattcaata 420
 ttttttatga attatttacc atttgtactt caatagtagc agaagacaaa aaaggtcatc 480
 taatacatgg gagaaacatg gattttggag tatttcttgg gtggaacata aataatgata 540
 cctgggtcat aactgagcaa ctaaacctt taacagtga tttggatttc caaagaaaca 600
 acaaaactgt cttcaaggct tcaagctttg ctggctatgt gggcatgtta acaggattca 660
 aaccaggact gttagctctt aactgaatg aacgtttcag tataaatggg gggtatctgg 720
 gtattctaga atggattctg ggaaagaaag atgccatgtg gatagggttc ctactagaa 780
 cagttctgga aaatagcaca agttatgaag aagccaagaa ttattgacc aagaccaaga 840
 tattggcccc agcctacttt atcctgggag gcaaccagtc tggggaagggt tgtgtgatta 900
 caccagacag aaaggaatca ttggatgtat atgaactcga tgctaagcag ggtagatggg 960

- 6 -

```

atgtggtaca aacaaattat gaccgttggg aacatccctt cttccttgat gatcgcagaa 1020
cgcttgcaaa gatgtgtctg aaccgcacca gccaaagagaa tatctcattt gaaaccatgt 1080
atgatgtcct gtcaacaaaa cctgtcctca acaagctgac cgtatacaca accttgatag 1140
atgttaccaa aggtcaattc gaaacttacc tgcgggactg ccctgaccct tgtatagggt 1200
ggtgagcaca cgtctggcct acagaatgcg gcctctgaga catgaagaca ccatctccat 1260
gtgaccgaac actgcagctg tctgacctc caaagactaa gactcgcggc aggttctctt 1320
tgagtcaata gcttgtcttc gtccatctgt tgacaaatga cagatctttt tttttttccc 1380
cctatcagtt gatttttctt atttacagat aacttcttta ggggaagtaa aacagtcac 1440
tagaattcac tgagttttgt ttcactttga catttgggga tctggtgggc agtcgaacca 1500
tggtgaactc cacctccgtg gaataaatgg agattcagcg tgggtgttga atccagcacg 1560
tctgtgtgag taacgggaca gtaaacactc cacattcttc agtttttcac ttctacctac 1620
atatttgtat gtttttctgt ataacagcct tttccttctg gttctaactg ctgttaaaat 1680
taatatatca ttatctttgc tgttattgac agcgatatta ttttattaca tatcattaga 1740
gggatgagac agacattcac ctgtatattt cttttaatgg gcacaaaatg ggcccttgcc 1800
tctaaatagc actttttggg gttcaagaag taatcagtat gcaaagcaat cttttataca 1860
ataattgaag tgttcccttt ttcataatta ctctacttcc cagtaaccct aaggaagttg 1920
ctaacttaaa aaactgcac ccacgttctg ttaatttagt aaataaacia gtcaaagact 1980
tgtggaaaat aggaagtga cccatatttt aaattctcat aagtagcatt gatgtaataa 2040
acaggttttt agtttgttct tcagattgat agggagtttt aaagaaattt tagtagttac 2100
taaaattatg ttactgtatt tttcagaaat caaactgctt atgaaaagta ctaatagaac 2160
ttgttaacct ttctaacctt cagattaac tgtgaaatgt acgtcatttg tgcaagaccg 2220
tttgtccact tcattttgta taatcacagt tgtgttctg acactcaata aacagtcact 2280
ggaaagagtg ccagtcagca gtcatgcacg ctgataaaaa aaaaaaaaaa aaa 2333

```

```

<210> 6
<211> 2530
<212> DNA
<213> Homo Sapiens

```

```

<400> 6
cagcttcctt gtggtttccc gaggttccct tgcttccgc tctgcgagga gcctttcatc 60
cgaaggcggg acgatgccg ataatcggca gccgaggaa cggcagccga ggatccgctc 120
cggaacgag cctcgttccg cggccgccat ggaaccggat ggtcgcgggtg cctgggcccc 180
cagtcgcgcc gcgctcgacc gcctggagaa gctgctgcgc tgctcgcgtt gtactaacat 240

```

tctgagagag cctgtgtgtt taggaggatg tgagcacatc ttctgtagta attgtgtaag	300
tgactgcatt ggaactggat gtccagtgtg ttacaccccg gcctggatac aagacttgaa	360
gataaataga caactggaca gcatgattca actttgtagt aagcttcgaa atttgctaca	420
tgacaatgag ctgtcagatt tgaaagaaga taaacctagg aaaagtttgt ttaatgatgc	480
aggaaacaag aagaattcaa ttaaaatgtg gtttagccct cgaagtaaga aagtcagata	540
tgttgtgagt aaagcttcag tgcaaaccga gcctgcaata aaaaaagatg caagtgtcga	600
gcaagactca tatgaatttg tttccccaag tcctcctgca gatgtttctg agagggctaa	660
aaaggcttct gcaagatctg gaaaaaagca aaaaaagaaa acttttagctg aatcaacca	720
aaaatggaat ttagaggcag aaaaagaaga tgggtgaattt gactccaaag aggaatctaa	780
gcaaaagctg gtatccttct gtagccaacc atctgttata tccagtcctc agataaatgg	840
tgaaatagac ttactagcaa gtggctcctt gacagaatct gaatgttttg gaagtttaac	900
tgaagtctct ttaccattgg ctgagcaaat agagtctcca gacactaaga gcaggaatga	960
agtagtgact cctgagaagg tctgcaaaaa ttatcttaca tctaagaaat ctttgccatt	1020
agaaaataat ggaaaacgtg gccatcacia tagactttcc agtcccattt ctaagagatg	1080
tagaaccagc attctgagca ccagtggaga ttttgtaag caaacctgac cctcagaaaa	1140
tataccattg cctgaatgtt cttcaccacc ttcattgcaa cgtaaagtgt gtggtacatc	1200
aggaggagaaa aacagtaaca tgtccgatga attcattagt ctttcaccag gtacaccacc	1260
ttctacatta agtagttcaa gttacaggca agtgatgtct agtccctcag caatgaagct	1320
gttgcccaat atggctgtga aaagaaatca tagaggagag actttgctcc atattgcttc	1380
tattaagggc gacatacctt ctgttgaata ctttttacia aatggaagtgt atccaaatgt	1440
taaagaccat gctggatgga caccattgca tgaagcttgc aatcatgggc acctgaaggt	1500
agtggaaatta ttgctccagc ataaggcatt ggtgaacacc accgggtatc aaaatgactc	1560
accacttcac gatgcagcca agaattggga cgtggatata gtcaagctgt tactttccta	1620
tggagcctcc agaaatgctg ttaatatatt tggctctgag cctgtcgatt atacagatga	1680
tgaaagtatg aaatcgctat tgctgtacc agagaagaat gaatcatcct cagctagcca	1740
ctgctcagta atgaacactg ggcagcgtag ggatggacct cttgtactta taggcagtgg	1800
gctgtcttca gaacaacaga aaatgctcag tgagcttgca gtaattctta aggctaaaaa	1860
atatactgag ttgacagta cagtaactca tgttgttgtt cctgggtgatg cagttcaaag	1920
taccttgaag tgtatgcttg ggattctcaa tggatgctgg attctaaaaat ttgaatgggt	1980
aaaagcatgt ctacgaagaa aagtatgtga acaggaagaa aagtatgaaa ttctgaagg	2040

- 8 -

tccacgcaga agcaggctca acagagaaca gctgttgcca aagctgtttg atggatgcta	2100
cttctatattg tggggaacct tcaaacacca tccaaaggac aaccttatta agctcgtcac	2160
tgcagggtggg ggccagatcc tcagtagaaa gcccaagcca gacagtgcg tgactcagac	2220
catcaataca gtcgcatacc atgcgagacc cgattctgat cagcgcttct gcacacagta	2280
tatcatctat gaagatttgt gtaattatca cccagagagg gttcggcagg gcaaagtctg	2340
gaaggctcct tcgagctggg ttatagactg tgtgatgtcc tttgagttgc ttcctcttga	2400
cagctgaata ttataccaga tgaacatttc aaattgaatt tgcacggttt gtgagagccc	2460
agtcattgta ctgttttttaa tgttcacatt tttacaaata ggtagagtca ttcataattg	2520
tctttgaatc	2530

<210> 7

<211> 1203

<212> DNA

<213> Homo Sapiens

<400> 7

ggacgctgat gcgtttgggt tctcgtctgc agaccctctg gacctggtca cgattccata	60
atgtaccaca acagtagtca gaagcggcac tggaccttct ccagcgagga gcagctggca	120
agactgcggg ctgacgcaa cgcgaaattc agatgcaaag ccgtggccaa cgggaagggt	180
cttccgaatg atccagtctt tcttgagcct catgaagaaa tgacactctg caaatactat	240
gagaaaagggt tattggaatt ctgttcggtg ttttaagccag caatgccaa atctgtttgtg	300
ggtacggctt gtatgtatctt caaacgtttt tatcttaata actcagtaat ggaatatcac	360
cccaggataa taatgctcac ttgtgcattt ttggcctgca aagtagatga attcaatgta	420
tctagtcctc agtttggttg aaacctccgg gagagtcctc ttggacagga gaaggcactt	480
gaacagatac tggaatatga actacttctt atacagcaac ttaatttcca ccttattgtc	540
cacaatcctt acagaccatt tgagggcttc ctcacgcact taaagaccg ctatcccata	600
ttggagaatc cagagatttt gaggaaaaca gctgatgact ttcttaatag aattgcattg	660
acggatgctt accttttata cacaccttcc caaattgccc tgactgccat tttatctagt	720
gcctccaggg ctggaattac tatggaaagt tatttatcag agagtctgat gctgaaagag	780
aacagaactt gcctgtcaca gttactagat ataatgaaaa gcatgagaaa cttagtaaag	840
aagtatgaac caccagatc tgaagaagtt gctgttctga aacagaagtt ggagcgatgt	900
cattctgctg agcttgcaact taacgtaatc acgaagaaga ggaaaggcta tgaagatgat	960
gattacgtct caaagaaatc caaacatgag gaggaagaat ggactgatga cgacctggtg	1020

- 9 -

gaatctctct aaccatttga agttgatttc tcaatgctaa ctaatcaaga gaagtaggaa 1080
 gcatatcaaa cgtttaactt tatttaaaaa gtataatgtg aaaacataaa atatattaaa 1140
 acttttctat tgttttcttt ccctttcaca gtaactttat gtaaaataaa ccattcttcaa 1200
 aag 1203

<210> 8
 <211> 653
 <212> DNA
 <213> Homo Sapiens

<400> 8
 atggcgctccc ttcccttgc acctgttaac atctttaagg caggagctga tgaagagaga 60
 gcagagacag ctctgtctgac ttcttttatt ggtgccatcg ccattggaga cttggtaaag 120
 agcaccttgg gacccaaagg catggacaaa attcttctaa gcagtggacg agatgcctct 180
 cttatggtaa ccaatgatgg tgccactatt ctaaaaaaca ttggtgttga caatccagca 240
 gctaaagttt tagttgatat gtcaagggtt caagatgatg aagttggtga tggcactacc 300
 tctgttaccg ttttagcagc agaattatta agggaagcag aatctttaat tgcaaaaaag 360
 attcatccac agaccatcat agcgggttgg agagaagcca cgaaggctgc aagagaggcg 420
 ctgttgagtt ctgcagttga tcatggttcc gatgaagtta aattccgtca agatttaatg 480
 aatattgcgg gcacaacatt atcctcaaaa cttcttactc atcacaaga ccactttaca 540
 aagttagctg tagaagcagt tctcagactg aaaggctctg gcaacctgga ggcaattcat 600
 attatcaaga agctaggagg aagtttgga gattcctatt tagatgaagg tat 653

<210> 9
 <211> 1686
 <212> DNA
 <213> Homo Sapiens

<400> 9
 ccacgcgtcc gggcgtaagc caggcggtgtt aaagccggtc ggaactgctc cggagggcac 60
 gggctccgta ggcaccaact gcaaggacce ctccccctgc gggcgctccc atggcacagt 120
 tcgcgttcga gagtgaacct cactcgctgc ttcagctgga tgcaccatc cccaatgcac 180
 cccctgcgcg ctggcagcgc aaagccaagg aagccgcagg cccggccccc tcaccatgc 240
 gggccgcaa cccatccac agcgcgggca ggactccggg ccgaactcct ggcaaatcca 300
 gttccaaggt tcagaccact cctagcaaac ctggcggtga ccgctatata ccccatcgca 360
 gtgctgccc gatggaggtg gccagcttcc tctgagcaa ggagaaccag tctgaaaaca 420
 gccagacgcc caccaagaag gaacatcaga aagcctgggc tttgaacctg aacggttttg 480

-10-

```

atgtagagga agccaagatc cttcggctca gtggaaaacc acaaaatgcg ccagaggggtt 540
atcagaacag actgaaagta ctctacagcc aaaaggccac tcctggctcc agccggaaga 600
cctgccgtta cattccttcc ctgccagacc gtatcctgga tgcgcctgaa atccgaaatg 660
actattacct gaaccttggtg gattggagtt ctgggaatgt actggccgtg gcaactggaca 720
acagtgtgta cctgtggagt gcaagctctg gtgacatcct gcagcttttg caaatggagc 780
agcctgggga atatatatcc tctgtggcct ggatcaaaga gggcaactac ttggctgtgg 840
gcaccagcag tgctgaggtg cagctatggg atgtgcagca gcagaaacgg cttcgaaata 900
tgaccagtca ctctgcccga gtgggctccc taagctggaa cagctatatc ctgtccagtg 960
gttcacgttc tggccacatc caccaccatg atgttcgggt agcagaacac catgtggcca 1020
cactgagtgg ccacagccag gaagtgtgtg ggctgcgctg ggccccagat ggacgacatt 1080
tggccagtgg tggtaatgat aacttgggtc atgtgtggcc tagtgctcct ggagaggggtg 1140
gctgggttcc tctgcagaca ttcaaccagc atcaaggggc tgtcaaggcc gtagcatggt 1200
gtccctggca gtccaatgtc ctggcaacag gagggggcac cagtgatega cacattcgca 1260
tctggaatgt gtgctctggg gcctgtctga gtgccgtgga tgccattcc caggtgtgct 1320
ccatcctctg gtctcccat tacaaggagc tcattctcagg ccatggcttt gcacagaacc 1380
agctagttat ttggaagtac ccaaccatgg ccaagggtggc tgaactcaaa ggtcacacat 1440
cccgggtcct gagtctgacc atgagcccag atggggccac agtggcatcc gcagcagcag 1500
atgagaccct gaggctatgg cgctgttttg agttggacce tgcgcggcgg cgggagcggg 1560
agaaggccag tgcagccaaa agcagcctca tccaccaagg catccgctga agaccaaccc 1620
atcacctcag ttgtttttta tttttctaataaagtcattgt ctcccttcat gttttttttt 1680
ttaaaa 1686

```

```

<210> 10
<211> 1374
<212> DNA
<213> Homo Sapiens

```

```

<400> 10
attgcggcgg cgccagagct gctggagcgc tcgggggtccc cgggcggcgg cggcggcgca 60
gaggaggagg caggcggcgg ccccggtggc tccccccggg acggtgcgcg gcccggcccg 120
tctcgcgaac tcgcggtggt cgcgcggccc cgcgtgctc cgaccccggg cccctccgcc 180
gccgccatgg ctgcggcgt agtgcccagc tcgcagaagg cgctgctgct ggagctcaag 240
gggctgcagg aagagccggt cgagggattc cgcgtgacac tgggtggacga gggcgatcta 300
tacaactggg aggtggccat tttcgggccc cccaacacct actacgaggg cggctacttc 360

```

- 11 -

aaggcgcgcc	tcaagttccc	catcgactac	ccatactctc	caccagcctt	tcggttcctg	420
accaagatgt	ggcaccctaa	catctacgag	acgggggacg	tgtgtatctc	catcctccac	480
ccgccggtgg	acgaccccca	gagcggggag	ctgccctcag	agaggtggaa	ccccacgcag	540
aacgtcagga	ccattctcct	gagtgtgatc	tccctcctga	acgagcccaa	caccttctcg	600
cccgcaaacg	tggacgcctc	cgtgatgtac	aggaagtgga	aagagagcaa	ggggaaggat	660
cgggagtaca	cagacatcat	ccggaagcag	gtcctgggga	ccaaggtgga	cgcgagcgt	720
gacggcgtga	aggtgcccac	cacgctggcc	gagtactgcg	tgaagaccaa	ggcgccggcg	780
cccgcagagg	gctcagacct	cttctacgac	gactactacg	aggacggcga	ggtggaggag	840
gaggccgaca	gctgcttcgg	ggacgatgag	gatgactctg	gcacggagga	gtcctgacac	900
caccagaata	aacttgccga	gtttacctca	ctagggccgg	accctgggct	ccttagacga	960
cagactacct	cacggagggt	ttgtgctggt	ccccgtctcc	tctggttggt	tcgttttggc	1020
tttttctccc	tccccatgtc	tggtctgggt	tttcaagtgc	ttcagagaag	aggggctgcc	1080
ccaccgccac	tcacgtcact	cggggctcgg	tggacgggcc	caggggtggga	gcggccggcc	1140
cacctgtccc	ctcgggaggg	gagctgagcc	cgacttctac	cgggggtcccc	cagcttccgg	1200
actggccgca	ccccggagga	gccacggggg	cgctgctggg	aacgtgggcg	gggggcccgtt	1260
tcctgacact	accagcctgg	gaggcccagg	tgtagcggtc	cgagggggccc	ggtcctgcct	1320
gtcagctcca	ggtcctggag	ccacgtccag	cactgagtgg	acggattcac	caat	1374

<210> 11
 <211> 806
 <212> DNA
 <213> Homo Sapiens

<400> 11	
cggcactggt	ctcgacgtgg ggcggccagc gatggagccg ccagttcaa tacaacaag 60
tgagtttgac	tcatcagatg aagagcctat tgaagatgaa cagactccaa ttcatatatc 120
atggctatct	ttgtcacgag tgaattgttc tcagtttctc ggtttatgtg ctcttccagg 180
ttgtaaat	aaagatgtta gaagaaatgt ccaaaaagat acagaagaac taaagagctg 240
tggtatacaa	gacatatttg ttttctgcac cagaggggaa ctgtcaaaat atagagtccc 300
aaaccttctg	gatctctacc agcaatgtgg aattatcacc catcatcatc caatcgcaga 360
tggagggact	cctgacatag ccagctgctg tgaaataatg gaagagctta caacctgcct 420
taaaaattac	cgaaaaacct taatacactg ctatggagga cttgggagat cttgtcttgt 480
agctgcttgt	ctcctactat acctgtctga cacaatatca ccagagcaag ccatagacag 540

- 12 -

cctgcgagac ctaagaggat ccggggcaat acagaccatc aagcaatata attatcttca	600
tgagtttcgg gacaaattag ctgcacatct atcatcaaga gattcacaat caagatctgt	660
atcaagataa aggaattcaa atagcatata tatgaccatg tctgaaatgt cagttctcta	720
gcataatttg tattgaaatg aaaccaccag tggtatcaac ttgaatgtaa atgtacatgt	780
gcagatattc ctaaagtttt attgac	806

<210> 12
 <211> 717
 <212> DNA
 <213> Homo Sapiens

<400> 12	
agagcgatca tgtcgcacaa acaaatttac tattcggaca aatacgacga cgaggagttt	60
gagtatcgac atgtcatgct gcccaaggac atagccaagc tggtccttaa aacccatctg	120
atgtctgaat ctgaatggag gaatcttggc gttcagcaga gtcagggatg ggtccattat	180
atgatccatg aaccagaacc tcacatcttg ctgttccggc gccactacc caagaaacca	240
aagaaatgaa gctggcaagc tacttttcag cctcaagctt tacacagctg tccttacttc	300
ctaacatctt tctgataaca ttattatggt gccttcttgt ttctcacttt gatattttaa	360
agatgttcaa tacactgttt gaatgtgctg gtaactgctt tgcttcttga gtagagccac	420
caccaccata gccagccag atgagtgtct tgtggacca cagcctaagc tgagtgtgac	480
cccagaagcc acgatgtgct ctgtatccag aacacacttg gcagatggag gaagcatctg	540
agtttgagac catggctgtt acagggatca tgtaaacttg ctgtttttgt tttttctgcc	600
gggtgttgta tgtgtggtga cttgcggtt tatgtttcag tgtactggaa actttccatt	660
ttattcaaga aatctgttca tgttaaaagc cttgattaaa gaggaagttt ttataat	717

<210> 13
 <211> 627
 <212> DNA
 <213> Homo Sapiens

<400> 13	
agtctccggc gagttgttgc ctgggctgga cgtggttttg tctgctgcgc ccgctcttcg	60
cgctctcggt tcattttctg cagcgcgcca cgaggatggc ccacaagcag atctactact	120
cggacaagta cttcgacgaa cactacgagt accggcatgt tatgttacct agagaacttt	180
ccaaacaagt acctaaaact catctgatgt ctgaagagga gtggaggaga cttggtgtcc	240
aacagagtct aggctgggtt cattacatga ttcatgagcc agaaccacat attcttctct	300
ttagacgacc ttttccaaaa gatcaacaaa aatgaagttt atctggggat cgtcaaatct	360

-13-

ttttcaaatt taatgtatat gtgtatataa ggtagtattc agtgaatact tgagaaatgt 420
 acaaatcttt catccatacc tgtgcatgag ctgtattctt cacagcaaca gagctcagtt 480
 aaatgcaact gcaagtaggt tactgtaaga tgtttaagat aaaagttctt ccagtcagtt 540
 tttctcttaa gtgcctgttt gagtttactg aaacagttta cttttgttca ataaagtttg 600
 tatgttgcac ttaaaaaaaaa aaaaaaa 627

<210> 14
 <211> 341
 <212> DNA
 <213> Homo Sapiens

<400> 14
 aggagaaggg aggtgactcc ggcggaagag gacaaggcag aatgcaggcc cttcggtgtg 60
 cccaggecgt gatccgctcc ttcagctcca ccgcccgaa ccgctttcag aaccgagtgc 120
 gcgagaaaca gaagctcttc caggaggaca atgacatccc gttgtacctg aagggcggca 180
 tcgttgacaa catcctgtac cgagtgacaa tgacgctgtg tctgggcggc actgtctaca 240
 gcttgacttc ccttggtctg gctccttcc ccaggaatta agaccaagaa gcctgggggg 300
 cctgagagac ttgaacaagt gtcaataaac gctggcctct g 341

<210> 15
 <211> 1581
 <212> DNA
 <213> Homo Sapiens

<400> 15
 ataactaaat tacatcttct tggctctttg actatgaaat agtttaccct agcaacatga 60
 aaaacaagag acctaagcta ttagaagaaa tgcagttcta tgtatcttgt gtgtatagtt 120
 tttccctggg tggttttcaa cgaccagtga ctcccttagct ggtttcctca gctgctagca 180
 cttgctctgg gtacttgtcc tcaacacgtc catctgcaac aatgtgtgcc taggaaataa 240
 actcaactta ctactacccc aacaaaaatg taatttttta aacgcagcac aactgggtg 300
 gattccaaag tcatgattat gctttactat gcactctgta ctattcagac cactactctc 360
 attcattact gcaattaaact gcacacataa ctatttttta ttgctaatta tacaccactg 420
 atttccactt taaaaaaaca ttagcatttg tctctaatta aatatttact gcttgtgttt 480
 tacagaccgg atatcaggtt cttctttaga ctgggcttat gacctgggca tcaaacacac 540
 atttgccttt gagctccgag ataaaggcaa atttggtttt ctcttccag aatcccggat 600
 aaagccaacg tgcagagaga ccattgctagc tgtcaaattt attgccaagt atactctcaa 660
 gcatacttcc taaagaactg ccctctgttt ggaataagcc aattaatcct tttttgtgcc 720

-14-

```

tttcatcaga aagtcaatct tcagttatcc ccaaatgcag cttctatttc acctgaatcc 780
ttctcttgct catttaagtc ccatgttact gctgtttgct tttacttact ttcagtagca 840
ccataacgaa gtagctttaa gtgaaacctt ttaactacct ttctttgctc caagtgaagt 900
ttggacccag cagaaagcat ttttttgaaa ggtgatatac agtggggcac agaaaacaaa 960
tgaaaaccct cagttttctca cagattttca ccatgtggct tcatcaattt atgtgctaatt 1020
acaataaaat aaaatgcact taatgcttta aaattcatct ttttatgata aacaatattc 1080
tctgtatttc tctatagcat taataatcaa tattaatgcc attcattcag tctgttaata 1140
agaaataata tcttcaattt tcaaaaacat aatttgcccta tctttttctg atagaagtag 1200
acattgttta tatcttcaaa aaagcaaaag gatgtcctag caggaaataa agtggttcat 1260
atagagatga atctcagtc tttaaataac cgatccagtt ctcatcagca taatgtacat 1320
taaattcaaa atagtttaaat ttaacctgcc ataatcagaa gaaaccacct gctaaaacat 1380
ctgtttgccc gtacagacac agacaagaca gtctggtcag ctgtgacccc tgccctccta 1440
atggatagaa aggaaacctg gaaacatact gtaagttgag gacggaaagt catgttgacc 1500
aaaggcaatc agggtaactt gctgcatttg taccatttat actcctatta tttaagatag 1560
tattattgga tagcttctcc c 1581

```

<210> 16
 <211> 2443
 <212> DNA
 <213> Homo Sapiens

```

<400> 16
aaatggcgtg cccgtctctc cgccggcccc ctgcctcgca gtggtttctc ctgcagctcc 60
cctgggctcc gcggccagta gtgcagcccg tggagccgcg gctttgcccc tctcctctgg 120
gtggcccccag tgcgcgggct gacactcatt cagccgggga aggtgaggcg agtagaggct 180
ggtgcggaac ttgccgcccc cagcagcgcc ggcgggctaa gccagggcc gggcagacaa 240
aagaggccgc ccgcgtagga aggcacggcc ggcggcgcg gagcgcagcg atggccgggc 300
gagggggcag cgcgctgctg gctctgtgag gggcactggc tgcttgcggg tggctcctgg 360
gcgccgaagc ccaggagccc ggggcgcccc cggcgggcat gaggcggcgc cggcggtgc 420
agcaagagga cggcatctcc ttcgagtacc accgctacce cgagctgcgc gaggcgctcg 480
tgtccgtgtg gctgcagtgc accgccatca gcaggattta cacgggtggg cgcagcttcg 540
agggccggga gtcctggtc atcgagctgt ccgacaacce tggcgtccat gagcctgggtg 600
agcctgaatt taaatacatt gggaatatgc atgggaatga ggctgttgga cgagaactgc 660
tcattttctt ggcccagtac ctatgcaacg aataccagaa ggggaacgag acaattgtca 720

```

- 15 -

acctgatcca	cagtacccgc	attcacatca	tgccttcct	gaaccagat	ggctttgaga	780
aggcagcgtc	tcagcctggt	gaactcaagg	actggtttgt	gggtcgaagc	aatgcccagg	840
gaatagatct	gaaccggaac	tttccagacc	tggataggat	agtgtacgtg	aatgagaaag	900
aaggtggtcc	aaataatcat	ctgttgaaaa	atatgaagaa	aattgtggat	caaaacacaa	960
agcttgctcc	tgagaccaag	gctgtcattc	attggattat	ggatattcct	tttgtgcttt	1020
ctgccaatct	ccatggagga	gaccttgtgg	ccaattatcc	atatgatgag	acgcgagtg	1080
gtagtgctca	cgaatacagc	tcctccccag	atgacgccat	tttccaaagc	ttggcccggg	1140
catactcttc	tttcaaccgc	gccatgtctg	acccaatcg	gccaccatgt	cgcaagaatg	1200
atgatgacag	cagctttgta	gatggaacca	ccaacggtgg	tgcttggtac	agcgtacctg	1260
gagggatgca	agacttcaat	taccttagca	gcaactgttt	tgagatcacc	gtggagctta	1320
gctgtgagaa	gttcccacct	gaagagactc	tgaagaccta	ctgggaggat	aacaaaaact	1380
ccctcattag	ctaccttgag	cagatacacc	gaggagttaa	aggatttgtc	cgagaccttc	1440
aaggtaaccc	aattgcgaat	gccaccatct	ccgtggaagg	aatagaccac	gatgttacat	1500
ccgcaaagga	tggtgattac	tggagattgc	ttatacctgg	aaactataaa	cttacagcct	1560
cagctccagg	ctatctggca	ataacaaaga	aagtggcagt	tccttacagc	cctgctgctg	1620
gggttgattt	tgaactggag	tcattttctg	aaaggaaaga	agaggagaag	gaagaattga	1680
tggaatggtg	gaaaatgatg	tcagaaactt	taaattttta	aaaaggcttc	tagttagctg	1740
ctttaaatct	atctatataa	tgtagtatga	tgtaatgtgg	tctttttttt	agattttgtg	1800
cagttaatac	ttaacattga	tttatttttt	aatcatttaa	atattaatca	actttcctta	1860
aaataaatag	cctcttaggt	aaaaatataa	gaacttgata	tatttcattc	tcttatatag	1920
tattcathtt	cctacctata	ttacacaaaa	aagtatagaa	aagatttaag	taattttgcc	1980
atcctaggct	taaatgcaat	attcctggta	ttattttaca	tgcagaattt	tttgagtaat	2040
tctagctttc	aaaaattagt	gaagttcttt	tactgtaatt	ggtgacaatg	tcacataatg	2100
aatgctattg	aaaaggttta	cagatacagc	tcggagttgt	gagcactcta	ctgcaagact	2160
taaatagttc	agtataaatt	gtcgtttttt	tcttgtgctg	actaactata	agcatgatct	2220
tgттаатгса	tttttgatgg	gaagaaaagg	tacatgttta	caaagagggt	ttatgaaaag	2280
aataaaaaatt	gacttcttgc	ttgtacatat	aggagcaata	ctattatatt	atgtagtccg	2340
ttaacactac	ttaaaagttt	agggttttct	cttggttgta	gagtggccca	gaattgcatt	2400
ctgaatgaat	aaagggttaa	aaaaaatccc	cagtgaaaaa	aaa		2443

-16-

<210> 17
 <211> 3100
 <212> DNA
 <213> Homo Sapiens

<400> 17

```

actcgtctct ggtaaagtct gagcaggaca ggggtggctga ctggcagatc cagaggttcc      60
cttggcagtc cacgccaggc cttcaccatg gatcagttcc ctgaatcagt gacagaaaac      120
tttgagtacg atgatttggc tgaggcctgt tatattgggg acatcgtggg ctttgggact      180
gtgttcctgt ccatattcta ctccgtcadc tttgccattg gcctgggtggg aaatttgttg      240
gtagtgtttg ccctcaccaa cagcaagaag cccaagagtg tcaccgacat ttacctcctg      300
aacctggcct tgtctgatct gctgtttgta gccactttgc ctttctggac tcactatttg      360
ataaatgaaa agggcctcca caatgccatg tgcaaattca ctaccgcctt cttcttcadc      420
ggcttttttg gaagcatatt cttcatcacc gtcacagca ttgataggta cctggccatc      480
gtcctggccg ccaactccat gaacaaccgg accgtgcagc atggcgtcac catcagccta      540
ggcgtctggg cagcagccat tttggtggca gcaccccagt tcatgttcac aaagcagaaa      600
gaaaatgaat gccttgggtga ctaccccgag gtcctccagg aaatctggcc cgtgctccgc      660
aatgtggaat caaattttct tggttctcta ctccccctgc tcattatgag ttattgctac      720
ttcagaatca tccagacgct gttttcctgc aagaaccaca agaaagccaa agccattaaa      780
ctgatccttc tgggtggcat cgtgtttttc ctcttctgga caccctacaa cgttatgatt      840
ttcctggaga cgcttaagct ctatgacttc tttcccagtt gtgacatgag gaaggatctg      900
aggctggccc tcagtgtgac tgagacggtt gcatttagcc attgttgccct gaatcctctc      960
atctatgcat ttgctgggga gaagttcaga agataccttt accacctgta tgggaaatgc     1020
ctggctgtcc tgtgtgggcg ctcagtcac gttgatttct cctcatctga atcacaaagg     1080
agcaggcatg gaagtgttct gagcagcaat tttacttacc acacgagtga tggagatgca     1140
ttgctccttc tctgaaggga atcccaaagc cttgtgtcta cagagaacct ggagttcctg     1200
aacctgatgc tgactagtga ggaaagattt ttgttgttat ttcttacagg cacaaaatga     1260
tggacccaat gcacacaaaa caaccctaga gtgttgttga gaattgtgct caaaatttga     1320
agaatgaaca aattgaactc tttgaatgac aaagagtaga catttctctt actgcaaagt     1380
tcacagaac tttttggttt gcagatgaca aaaattcaac tcagactagt ttagttaaat     1440
gaggggtggg aatattgttc atattgtggc acaagcaaaa ggggtgtctga gccctcaaag     1500
tgaggggaaa ccagggcctg agccaagcta gaattccctc tctctgactc tcaaattctt     1560
tagtcattat agatccccc aactttacat gacacagctt tatcaccaga gagggactga     1620

```

-17-

caccatggtt tctctggccc caagggaaaa ttcccagga agtgctctga taggccaagt 1680
 ttgtatcagg tgcccatccc tggaagggtgc tgttatccat ggggaaggga tatataagat 1740
 ggaagcttcc agtccaatct catggagaag cagaaatata tatttccaag aagttggatg 1800
 ggtgggtact attctgatta cacaaaacaa atgccacaca tcacccttac catgtgacctg 1860
 atccagcctc tcccctgatt acaccagcct cgtcttcatt aagccctctt ccatcatgtc 1920
 cccaaacctg caaggggtcc ccaactgccta ctgcatcgag tcaaaactca aatgcttggc 1980
 ttctcatagc tccaccatgg ggtcctacca atagattccc cattgcctcc tcttcccaa 2040
 aggactccac ccatcctatc agcctgtctc ttccatatga cctcatgcat ctccacctgc 2100
 tcccaggcca gtaagggaaa tagaaaaacc ctgcccccaa ataagaaggg atggattcca 2160
 accccaactc cagtagcttg ggacaaatca agcttcagtt tcttggctctg tagaagaggg 2220
 ataagggtacc ttccacatag agatcatcct ttccagcatg aggaactagc caccaactct 2280
 tgcagggtctc aacccttttg tctgctctt agacttctgc tttccacacc tgcactgctg 2340
 tgctgtgccc aagttgtggt gctgacaaag cttggaagag cctgcagggtg ccttggccgc 2400
 gtgcatagcc cagacacaga agaggctggt tcttacgatg gcaccagtg agcactccca 2460
 agtctacaga gtgatagcct tccgtaacct aactctctg gactgccttg aatatcccct 2520
 cccagtcacc ttgtgcaagc ccctgcccct ctgggaaaat accccatcat tcatgctact 2580
 gccaacctgg ggagccaggg ctatgggagc agcttttttt tccccctag aaacgtttgg 2640
 aacaatgtaa aacttttaaag ctcgaaaaca attgtaataa tgctaaagaa aaagtcatcc 2700
 aatctaacca catcaatatt gtcattcctg tattcacccg tccagacctt gttcacactc 2760
 tcacatgttt agagttgcaa tcgtaatgta cagatggttt tataatctga tttgttttcc 2820
 tcttaacggt agaccacaaa tagtgctcgc tttctatgta gtttggtaat tatcatttta 2880
 gaagactcta ccagactgtg tattcattga agtcagatgt ggtaactgtt aaattgctgt 2940
 gtatctgata gctctttggc agtctatatg tttgtataat gaatgagaga ataagtcatg 3000
 ttccttcaag atcatgtacc ccaatttact tgccattact caattgataa acatttaact 3060
 tgtttccaat gtttagcaaa tacatatttt atagaacttc 3100

<210> 18
 <211> 3995
 <212> DNA
 <213> Homo Sapiens

<400> 18
 ggatccgcgg gacagatgag gaaggggctt aagtcactgc agccagaggg atggagggtg 60
 actgatggga gggcttctcc ggtgggggtta gaagggaaaa gtagggaaag agaagtgtaa 120

- 18 -

ggtagatggc agaggcagag acatggaaag acagactcta gggttcctga tgatatctat	180
ctcgccaac acaaaagga gggtagagt gtgggggcac ccaagctagg gtgtgagtac	240
cctaagtgt ttttctgag atgtaggcca ttcactaact cttggaacag ctacagtttc	300
acagtaggaa gacccccca gattcactgc ccctccctta gtaaagcctc tgagaccttc	360
ctgaacattc ctttctgtct ttgccctctg ttccttcag agactatgtg ccagggcaga	420
tggattcctc ccgggcctga gaggaactgc aggaattctc ctgcctctta cccgtaaaac	480
cccaacttct ctagccctag ggcaggaagt cccaacaat ttctaccct ttttctgcaa	540
ttctcattgg ggtgagagga ggcccaggag gagagagagc tgggctcagc ttctttttga	600
gctgctggag ccctctgtga ggaggccctc tttgctggct tctcaggaga gtgtggctag	660
gttctgctg cctatgggaa gagggggcca ggggtgtgtg agcaagatgg tgcggtgctg	720
gtgccttggg acctggggga atgggacagc tggctggctc agagacggcc tactttactc	780
acagctggaa tttagtgggg agaagcagct caactccaat cctggaggat tagggagatt	840
aaagtgagag aagagagaga tgtcccagag accaagagct ccaggtcag ccctctggct	900
cctggcacc cactgctgc ggtgggcacc ccactcctc acagtgtgc atagcgacct	960
cttcaggcc ttgctggaca tcctggacta ttatgaggct tcctctcag agagtcagaa	1020
ataccgctac caagatgaag acacgcccc tctggagcac agccggccc acctcccaa	1080
ccaggccaat tctccccag tgattgtcaa cacagatacc ctagaagccc caggatatga	1140
gttgcagggt aacgggaccg agggggagat ggaatacgag gaaatcacat tggaaagggg	1200
taactcaggt ctgggcttca gcatcgagg tggcactgac aaccacaca tcggtgacga	1260
cccatccatt ttcacacca agatcattcc tgggtgggct gcggcccagg atggccgct	1320
cagggtcaac gacagcatcc tgtttgtaa tgaagtggac gtgcgagagg tgaccactc	1380
agcggcgggt gaagccctca aagaggcagg ctccatcgtt cgcctctatg tcatgcgccg	1440
gaagccccg gctgagaagg tcatggagat caagctcatc aaggggccta aaggtcttgg	1500
cttcagcatc gcagggggcg tagggaacca gcacatccca ggagataata gcatctatgt	1560
aacaagatc atcgaagggg gtgctgcca caaggatggg aggttgaga ttggagacaa	1620
gatcctggcg gtcaacagt tggggctaga ggacgtcatg catgaagatg ctgtggcagc	1680
cctgaagaac acgtatgat ttgtctacct aaagtgggc aagcccagca atgcctacct	1740
gagtgcagc tatgctcccc cagacatcac aacctcttat tcccagcacc tggacaatga	1800
gatcagtcac agcagctacc tgggcaccga ctaccacaca gccatgacct ccacttcccc	1860
tcggcgctac tctccagtgg ccaaggacct gctcggggag gaagacattc cccgagaacc	1920

gaggcgaatt	gtgateccacc	ggggctccac	gggcctgggc	ttcaacatcg	tgggtggcga	1980
ggacggtgaa	ggcatcttca	tctcctttat	cctggccggg	ggccctgcag	acctcagtgg	2040
ggagctgcgg	aagggggacc	agatcctgtc	ggtcaacggt	gtggacctcc	gaaatgccag	2100
ccatgagcag	gctgccattg	ccctgaagaa	tgcgggtcag	acggtcacga	tcacgcctca	2160
gtataaacca	gaagagtaca	gccgattcga	ggccaagatc	cacgaccttc	gggaacagct	2220
catgaacagc	agcctgggct	cagggactgc	gtccttgctg	agcaacccca	aaaggggttt	2280
ctacatcagg	gccctgtttg	attacgacaa	gaccaaggac	tgcggcttcc	tgagccaggc	2340
cctgagcttc	cgctttgggg	atgtgctgca	tgtcatcgat	gctagtgatg	aggagtgggtg	2400
gcaggcacgg	cgggtccact	ctgacagtga	gaccgacgac	attgggttca	tccccagcaa	2460
acggcggggt	gagcgacgag	agtgggtcaag	gttaaaggcc	aaggactggg	gctccagctc	2520
tggatcgcat	ggtcgagaag	actcggttct	gagctacgag	acagtgacgc	agatggaagt	2580
gcactatgct	cgccccatca	tcctccttgg	gccaccaag	gaccgcgcca	acgatgatct	2640
tctctccgag	ttccccgaca	agtttggatc	ctgtgttccc	catacgacac	ggccaagcg	2700
ggagtatgag	atagatggcc	gggattacca	ctttgtgtcg	tcccgggaga	aaatggagaa	2760
ggacattcag	gcgacaagt	tcattgaggc	cggccagtac	aacagccacc	tctatgggac	2820
cagcgtccag	tccgtgagag	aggtggcaga	gcaggggaag	cactgcatcc	tcgatgtctc	2880
ggccaatgcc	gtgcggcggc	tgcaggcggc	ccacctgcac	cccatcgcca	tcttcatccg	2940
cccccgctcc	ctggagaatg	tgctagagat	taacaagcgg	atcacagagg	agcaagcccg	3000
caaagccttc	gacagagcca	ccaagctgga	gcaggagttc	acagagtgtc	tctcagccat	3060
cgtggaggggt	gacagctttg	aggagatcta	ccacaagggt	aagcgtgtca	tcgaggacct	3120
ctcaggcccc	tacatctggg	ttccagcccc	agagagactc	tgattcctgc	cctggcttgg	3180
cctggactcg	ccctgcctcc	atcacctggg	cccttgggtc	ggactgaatt	gccccagccc	3240
ttggctcccc	ccggcctccc	tcccaccctt	tcttatttat	ttcctttcta	actggatcca	3300
gcctgttggg	ggggggacac	tcctctgcat	gtatccccgc	accccagaac	tgggctcctg	3360
aacgccagga	acctgggggtc	tgggggggag	ctgggctcct	tgttccgagc	ccttgcctct	3420
taggatcccc	gccccacct	gccccaatg	cacacacaga	cccaccgggg	gccacctgcc	3480
ctcccccatc	ctctcccaca	cacattccag	aagtcagggc	cccctcgagg	agcaccgct	3540
gcagggatgc	agggccacag	gcctccgctc	tctcctaagg	caggggtctgg	ggtcacccct	3600
gcctcatcgt	aattccccat	gttaccttga	tttctcattt	attttttcca	ctttttttct	3660
tctcaaagggt	ggtttttttg	ggggagaagc	aggggactcc	gcagcggggc	cctgccttcc	3720

-20-

```

acatgcccc accatttttc tttgccggtt tgcattgagt gaaggtctaa atgtggcttt 3780
tttttttttt ttcctgggaa tttttttggg gaaaaggag ggatgggtct agggagtggg 3840
aaatgcggga gggaggggtg ggcaggggtc gggggtcggg tgtccgggag ccaggaaga 3900
ctggaaatgc tgccgccttc tgcaatttat ttattttttt cttttgagag agtgaaagga 3960
agagacagat acttgaaaaa aaaaaaaaaa aaaaa 3995

```

```

<210> 19
<211> 3025
<212> DNA
<213> Homo Sapiens

```

```

<400> 19
gcacgagcag gcagttcaga ttaaagaagc taattgatca agaaatcaag tctcaggagg 60
agaaggagca agaaaaggag aaaagggtca ccaccctgaa agaggagctg accaagctga 120
agtcttttgc tttgatggtg gtggatgaac agcaaaggct gacggcacag ctcacccttc 180
aaagacagaa aatccaagag ctgaccacaa atgcaaagga aacacatacc aaactagccc 240
ttgctgaagc cagagttcag gaggaagagc agaaggcaac cagactagag aaggaactgc 300
aaacgcagac cacaaagttt caccaagacc aagacacaat tatggcgaag ctcaccaatg 360
aggacagtca aaatcgccag cttcaacaaa agctggcagc actcagccgg cagattgatg 420
agttagaaga gacaaacagg tctttacgaa aagcagaaga ggagctgcaa gatataaaag 480
aaaaaatcag taaggggagaa tatggaaacg ctggtatcat ggctgaagtg gaagagctca 540
taaaaatgga ggagcagtg agagatctca ataagaggct tgaaagggag acgttacaga 600
gtaaagactt taaactagag gttgaaaaac tcagtaaaag aattatggct ctggaaaagt 660
tagaagacgc tttcaacaaa agcaaacaag aatgctactc tctgaaatgc aatttagaaa 720
aagaaaggat gaccacaaag cagttgtctc aagaactgga gagtttaaaa gtaaggatca 780
aagagctaga agccattgaa agtcggctag aaaagacaga attcactcta aaagaggatt 840
taactaaact gaaaacatta actgtgatgt ttgtagatga acggaaaaca atgagtgaag 900
aattaaagaa aactgaagat aaattacaag ctgcttcttc tcagcttcaa gtggagcaaa 960
ataaagtaac aacagttact gagaagttaa ttgaggaaac taaaagggcg ctcaagtcca 1020
aaaccgatgt agaagaaaag atgtacagcg taaccaagga gagagatgat ttaaaaaaca 1080
aattgaaagc ggaagaagag aaaggaaatg atctcctgtc aagagttaat atgttgaaaa 1140
ataggcttca atcattggaa gcaattgaga aagatttcct aaaaaacaaa ttaaatcaag 1200
actctgggaa atccacaaca gcattacacc aagaaaacaa taagattaag gagctctctc 1260

```

-21-

aagaagtgga	aagactgaaa	ctgaagctaa	aggacatgaa	agccattgag	gatgacctca	1320
tgaaaacaga	agatgaatat	gagactctag	aacgaaggta	tgctaatgaa	cgagacaaag	1380
ctcaattttt	atctaaagag	ctagaacatg	ttaaaatgga	acttgctaag	tacaagttag	1440
cagaaaagac	agagaccagc	catgaacaat	ggctttttcaa	aaggcttcaa	gaagaagaag	1500
ctaagtcagg	gcacctctca	agagaagtgg	atgcattaaa	agagaaaatt	catgaataca	1560
tggcaactga	agacctaata	tgtcacctcc	agggagatca	ctcagtctgc	aaaaaaaaac	1620
taaatcaaca	agaaaacagg	aacagagatt	taggaagaga	gattgaaaac	ctcactaagg	1680
agttagagag	gtaccggcat	ttcagtaaga	gcctcaggcc	tagtctcaat	ggaagaagaa	1740
tttccgatcc	tcaagtattt	tctaaagaag	ttcagacaga	agcagtagac	aatgaaccac	1800
ctgattacaa	gagcctcatt	cctctggaac	gtgcagtcac	caatgggtcag	ttatatgagg	1860
agagtgagaa	tcaagacgag	gaccctaattg	atgagggatc	tgtgctgtcc	ttcaaattgca	1920
gccagtctac	tccatgtcct	gttaacagaa	agctatggat	tccctggatg	aaatccaagg	1980
aggggcatct	tcagaatgga	aaaatgcaaa	ctaaacccaa	tgccaacttt	gtgcaacctg	2040
gagatctagt	cctaagccac	acacctgggc	agccacttca	tataaagggt	actccagacc	2100
atgtacaaaa	cacagccact	cttgaaatca	caagtccaac	cacagagagt	cctcactctt	2160
acacgagtac	tgcatgata	ccgaactgtg	gcacgccaaa	gcaaaggata	accatcctcc	2220
aaaacgcctc	cataacacca	gtaaagtcca	aaacctctac	cgaagacctc	atgaatttag	2280
aacaaggcat	gtccccaatt	accatggcaa	cctttgccag	agcacagacc	ccagagtctt	2340
gtggttctct	aactccagaa	aggacaatgt	ccctattcag	gttttggctg	tgactgggtc	2400
agctagctct	cctgagcagg	gacgtcccc	agaaccaaca	gaaatcagtg	ccaagcatgc	2460
gatattcaga	gtctccccag	accggcagtc	atcatggcag	tttcagcggt	caaacagcaa	2520
tagctcaagt	gtgataacta	ctgaggataa	taaaatccac	attcacttag	gaagtcctta	2580
catgcaagct	gtagccagcc	cttcagcacc	actgcaggat	aaccgaactc	aaggcttaat	2640
taacggggca	ctaaacaaaa	caaccaataa	agtcaccagc	agtattacta	tcacaccaac	2700
agccacacct	cttcctcgac	aatcacaaat	tacagtaagt	aatatatata	actgaccacg	2760
ctcacctcca	tccagtccat	actgatattt	ttgcaaggaa	ctcaatcett	ttttaatcat	2820
ccctccatat	cccccaagac	tgactgaact	cgtactttgg	gaaggtttgt	gcatgaacta	2880
tacaagagta	tctgaaacta	actgttgcct	gcatagtcac	atcgagtgtg	cacttactgt	2940
atatcttttc	atttacatac	ttgtatggaa	aatatttagt	ctgcacttgt	ataaatacat	3000
ctttatgtat	ttgaaaaaaa	aaaaa				3025

-22-

<210> 20
 <211> 599
 <212> DNA
 <213> Homo Sapiens

<400> 20
 cgggacgcgg atgcagacgc aggcggaggg gctgacggcg gggatggccg ggggtggccac 60
 agctgccgcg ggggcgtgga cacagccgca gctccggccg gtggagctcc cccagcgcac 120
 gcgccaggtc cgggcagaga cgccgcgtct gccgcagggg gtcacgaatg cggccgcaca 180
 tattcaccct cagcgtgcct ttccccaccc ccttggaggc ggaaatcgcc catgggtccc 240
 tggcaccaga tgccgagccc caccaaaggg tggttgggaa ggatctcaca gtgagtggca 300
 ggatcctggg cgtccgctgg aaagctgaag actgtcgcct gctccgaatt tccgtcatca 360
 actttcttga ccagctttcc ctggtgggtgc ggaccatgca gcgctttggg cccccgttt 420
 cccgctaagc ctggcctggg caaatggagc gaggtccac ttgctgtctc cttgtaggca 480
 gtgcgtccat ccttccctag ggcaggaatt cccacagttg ctactttcct gggagggcct 540
 catgttttat ctggttctta aatgtttggt actacagaaa ataaaactga ggtattatt 599

<210> 21
 <211> 890
 <212> DNA
 <213> Homo Sapiens

<400> 21
 ggcggaccga agaacgcagg aagggggccg gggggacccg ccccgggccg gccgcagcca 60
 tgaactccaa cgtggagaac ctacccccgc acatcatccg cctggtgtac aaggaggtga 120
 cgacactgac cgcagaccca cccgatggca tcaaggtctt tcccaacgag gaggacctca 180
 ccgacctcca ggtcaccatc gagggccctg aggggacccc atatgctgga ggtctgttcc 240
 gcatgaaact cctgctgggg aaggacttcc ctgcctcccc acccaagggc tacttcctga 300
 ccaagatctt ccacccgaac gtgggcgcca atggcgagat ctgcgtcaac gtgctcaaga 360
 gggactggac ggctgagctg ggcacccgac acgtactgct gaccatcaag tgctgtgta 420
 tccaccctaa ccccgagtct gcaactcaac aggaggcggg ccgctgtctc ttggagaact 480
 acgaggagta tgcggctcgg gcccgctctgc tcacagagat ccacgggggc gccggcgggc 540
 ccagcggcag ggccgaagcc ggtcgggccc tggccagtgg cactgaagct tcctccaccg 600
 accctggggc cccagggggc ccgggagggg ctgaggggtcc catggccaag aagcatgctg 660
 gcgagcgca taagaagctg gcggccaaga aaaagacgga caagaagcgg gcgctgcggg 720
 cgctgcggcg gctgtagtgg gctctcttcc tccttcacc gtgaccccaa cctctcctgt 780

-23-

ccccccctc caactctgtc tctaagttat ttaaattatg gctgggggtcg gggagggtac 840

agggggcact gggacctgga ttgtttttc taaataaagt tggaaaagca 890

<210> 22
 <211> 1449
 <212> DNA
 <213> Homo Sapiens

<220>
 <221> Unsure
 <222> (1316)..(1316)
 <223> n = a, c, g, or t

<220>
 <221> Unsure
 <222> (1360)..(1360)
 <223> n = a, c, g, or t

<220>
 <221> Unsure
 <222> (1366)..(1367)
 <223> n = a, c, g, or t

<220>
 <221> Unsure
 <222> (1369)..(1369)
 <223> n = a, c, g, or t

<400> 22
 agccgaaact gagagggggc ggactcacag tgatgtgcac ctctcccgt ccagggtggg 60
 cctgcctggg gaaagcttgt ggccggaaga gaaaatgagc ttctaggac ccctgactca 120
 cgacctcatc aacgttggtg ctactgcttg gtggagaatg taaacccttt gtaaccccat 180
 cccatgcccc tccgactccc caccagga gggaacgggc aggcggggcg gccttgacaga 240
 tccacagggc aaggaaacaa gaggggagcg gccaaagtgc cggaccagga ggccccctac 300
 ttcagaggca agggccatgt ggtcctggcc cccacccca tcccttccca cctaggagct 360
 cccctccac acagcctcca tctccagggg aacttggtgc tacacgttg tgctcttacc 420
 ttctggggg gagggaggag ggaaggggtg ccctcgggg aacccctac ctggggctcc 480
 tctaaagatg gtgcagacac ttctgggca gtccagctc cccctgcccc ccaggacca 540
 ccgttggtg ccatccagtt ggtacccaag cacctgaagc ctcaaagctg gattcgctct 600
 agcatccctc ctctctggg tccacttggc cgtctctcc ccaccgatcg ctgttcccca 660
 catctggggc gcttttgggt tggaaaacca cccacactg ggaatagcca ccttgccct 720
 tgtagaatcc atccgcgat ccgtccatc atccatcgg ccgtccatcc atgtccccag 780

-24-

ttgaccgccc ggcaccatta gctggctggg tgcaccacc atcaacctgg ttgacctgtc	840
atggccgcct gtgccctgcc tccaccccc tccctacactc cccagggcg tgcggggctg	900
tgcagactgg ggtgccaggc atctcctccc caccgggggt gtccccacat gcagtactgt	960
atacccccca tccctccctc ggtccactga acttcagagc agttcccatt cctgccccgc	1020
ccatcttttt gtgtctcgct gtgatagatc aataaatatt ttattttttg tccctggatat	1080
ttggggatta tttttgattg ttgatattct cttttggttt tattgttgtg gttcattgaa	1140
aaaaaaagat aatttttttt tctgatccgg ggagctgtat cccagtaga aaaaacattt	1200
taatcactct aatataactc tggatgaaac acaccttttt ttttaataag aaaagagaat	1260
taactgcttc agaaatgact aataaatgaa aaccctttaa aggaaactgt gtcttngctt	1320
ccttggtatg atttaatctg ctttcaactg ttggcctggg tggggnnang ggctctgctt	1380
cagggaacct ccaccacca aattgtatth gagaggttgc ccaacaaaa gccctgctg	1440
cctggcttc	1449

<210> 23
 <211> 736
 <212> DNA
 <213> Homo Sapiens

<400> 23	
cgagctggag aggtggctcg agaagtagga acctcctgcc gggctcgtgg cggcttctgt	60
ccgctccgcg gagggaagcg ccttccccac aggacatcaa tgcaagcttg aataagaaaa	120
acaaattctt cctcctaagc catggcatat cagttataca gaaatactac tttgggaaac	180
agtcttcagg agagcctaga tgagctcata cagtctcaac agatcacccc ccaacttgcc	240
cttcaagttc tacttcagtt tgataaggct ataaatgcag cactggctca gagggtcagg	300
aacagagtca atttcagggg ctctctaaat acgtacagat tctgcgataa tgtgtggact	360
tttgtactga atgatgttga attcagagag gtgacagaac ttattaaagt ggataaagtg	420
aaaattgtag cctgtgatgg taaaaatact ggctccaata ctacagaatg aatagaaaaa	480
atatgacttt ttacaccat cttctgttat tcattgcttt tgaagagaag catagaagag	540
actttttatt tattctagaa ttgcagaaat gactacactg tgctatacca gagaattcca	600
gtagaaagaa acttgtaact ctgtagcctc ttacatcacc ttattatac agcatgaaaa	660
accataactt ttttttaagg aaaaagttg ttgccttcct aagaaccttc ttttaataaac	720
tcattttaaa actctg	736

<210> 24

-25-

<211> 2212

<212> DNA

<213> Homo Sapiens

<400> 24

tgccggctgc	tcctcgacca	ggcctccttc	tcaacctcag	cccgcggcgc	cgacccttcc	60
ggcaccctcc	cgccccgtct	cgtactgtcg	ccgtcacgcg	cgcggtcccg	gccctggccc	120
cgatggctct	gtgcaacgga	gactccaagc	tggagaatgc	tggaggagac	cttaaggatg	180
gccaccacca	ctatgaagga	gctgttgtca	ttctggatgc	tggtgctcag	tacgggaaag	240
tcatagaccg	aagagtgagg	gaactgttcg	tgcagtctga	aattttcccc	ttggaaacac	300
cagcatttgc	tataaaggaa	caaggattcc	gtgctattat	catctctgga	ggacctaat	360
ctgtgtatgc	tgaagatgct	ccctggtttg	atccagcaat	attcactatt	ggcaagcctg	420
ttcttggaat	ttgctatggg	atgcagatga	tgaataaggt	atttgagggt	actgtgcaca	480
aaaaaagtgt	cagagaagat	ggagttttca	acattagtgt	ggataataca	tgttcattat	540
tcaggggcct	tcagaaggaa	gaagtgtgtt	tgcttacaca	tggagatagt	gtagacaaag	600
tagctgatgg	attcaagggt	gtggcacggt	ctggaaacat	agtagcaggc	atagcaaagt	660
aatctaaaaa	gttatatgga	gcacagttcc	accctgaagt	tggccttaca	gaaaatggaa	720
aagtaatact	gaagaatttc	ctttatgata	tagctggatg	cagtgggaacc	ttcacctgac	780
agaacagaga	acttgagtgt	attcgagaga	tcaaagagag	agtaggcacg	tcaaagtttt	840
tggttttact	cagtgggtga	gtagactcaa	cagtttgtac	agctttgcta	aatcgtgctt	900
tgaaccaaga	acaagtcatt	gctgtgcaca	ttgataatgg	ctttatgaga	aaacgagaaa	960
gccagtctgt	tgaagaggcc	ctcaaaaagc	ttggaattca	ggtcaaagtg	ataaatgctg	1020
ctcattcttt	ctacaatgga	acaacaaccc	taccaatatc	agatgaagat	agaaccccac	1080
ggaaaagaat	tagcaaaacg	ttaaataaga	ccacaagtcc	tgaagagaaa	agaaaaatca	1140
ttggggatac	ttttgttaag	attgccaatg	aagtaattgg	agaaatgaac	ttgaaaccag	1200
aggaggtttt	ccttgcccaa	ggtactttac	ggcctgatct	aattgaaagt	gcaccccttg	1260
ttgcaagtgg	caaagctgaa	ctcatcaaaa	cccatcacia	tgacacagag	ctcatcagaa	1320
agttgagaga	ggagggaaaa	gtaatagaac	ctctgaaaga	ttttcataaa	gatgaagtga	1380
gaattttggg	cagagaactt	ggacttccag	aagagttagt	ttccaggcat	ccatttccag	1440
gtcctggcct	ggcaatcaga	gtaatatgtg	ctgaagaacc	ttatatttgt	aaggactttc	1500
ctgaaaccaa	caatattttg	aaaatagtag	ctgatttttc	tgcaagtgtt	aaaaagccac	1560
ataccctatt	acagagagtc	aaagcctgca	caacagaaga	ggatcaggag	aagctgatgc	1620
aaattaccag	tctgcattca	ctgaatgcct	tcttgctgcc	aattaaaact	gtaggtgtgc	1680

-26-

```

agggtgactg tcgttcctac agttacgtgt gtggaatctc cagtaaagat gaacctgact 1740
gggaatcact tattttttctg gctaggetta tacctcgcat gtgtcacaac gttaacagag 1800
ttgtttatat atttgcccca ccagttaaag aacctcctac agatgttact cccactttct 1860
tgacaacagg ggtgctcagt actttacgcc aagctgattt tgaggcccat aacattctca 1920
gggagtctgg gtatgctggg aaaatcagcc agatgccggt gattttgaca ccattacatt 1980
ttgatcgga cccacttcaa aagcagcctt catgccagag atctgtgggt attcgaacct 2040
ttattactag tgacttcatg actggtatac ctgcaacacc tggcaatgag atccctgtag 2100
agggtgtatt aaagatggtc actgagatta agaagattcc tggattttct cgaattatgt 2160
atgacttaac atcaaagccc ccaggaacta ctgagtggga gtaataaact tc 2212

```

```

<210> 25
<211> 1585
<212> DNA
<213> Homo Sapiens

```

```

<400> 25
acagcagtta cactgcggcg ggcgtctgtt ctagtgtttg agccgtcgtg cttcaccggt 60
ctacctcgt agcatgtcgg gccgcggcaa gactggcggc aaggcccgcg ccaaggccaa 120
gtcgcgctcg tcgcgcgccg gcctccagtt ccagtgggc cgtgtacacc ggctgctgcg 180
gaagggccac tacgccgagc gcgttggcgc cggcgcgcca gtgtacctgg cggcagtgct 240
ggagtacctc accgctgaga tcctggagct ggcgggcaat gcggcccgcg acaacaagaa 300
gacgcgaatc atccccgcgc acctgcagct ggccatccgc aacgacgagg agctcaacaa 360
gctgctgggc ggcgtgacga tcgcccaggg aggcgtcctg cccaacatcc aggccgtgct 420
gctgcccaag aagaccagcg ccaccgtggg gccgaaggcg cctcggggcg gcaagaaggc 480
caccaggcc tcccaggagt actaagaggg ccgcgcgccg ggccggccgc cccagctccc 540
catgccacca caaaggccct tttaagggcc accaccgccc tcatggaaag agctgagccg 600
cttcagactg cggggcaagc gggccgcggc tcccttcccc tcccctcccc tcgcccgcct 660
tcgccgcccg gcctcgagtc ccgcgccgc cccgtccccg tcccgcaccg cctgccgcgt 720
cggcctcggg cctgcctgt ccgcgctccg cctccggta gggttcgggc cttccggatg 780
cggtttgggc gctcttcggg gacctccgtg gcgcggaaga cccgagcctg ccgggggggag 840
gccggcggcg ccgcacctgc ccgcctcggc gttcgtgact cagccgcccc atcccagtc 900
gctaaggggc tgcggggagg ccgcagcacc ttctggaaga cttggccttc cgctctgacg 960
cagggccgag gtgggcagtc caggccgaga gccggcggcc ctgaaggtga gtgaggccct 1020

```

-27-

```

cggcagctgc agccggggtg tctggtaccc ccccggcgtg gtgcttagcc caggactttc 1080
agacggccgc tggccgggag gctttggtgg gagagacgcg atcgccgatt tcggtctggc 1140
gccctttctg cggccgggac ccaggccttt cacatcagct ctccctccat cttcattcat 1200
aggtctgcgc tggggccggg acgaagcact tggtaacagg cacatcttcc tcccagagtga 1260
ctgcctccta ggaggacatt taggggaggg cagaggcctg cagtttggtt tcacggctgg 1320
ctatgtggac agcaagagtc gttttgcgga acgcgactgg cagccaggcc tgtcggggccc 1380
ccgacgccgc cccatttccc ttccagcaaa ctcaactcgg caatccaagc acctagatac 1440
cagcacaagt cggttaatcc ctgtctggac tgagcctccg ttggcttctg aactggaatt 1500
ctgcagctaa cccttccacg actagaacct taggcattgg ggagtttttag atggactaat 1560
tttattaaag gattgttttt ttttt 1585

```

<210> 26
 <211> 847
 <212> DNA
 <213> Homo Sapiens

```

<400> 26
agtggcttcc taacagcaga agaactaaca atccactgaa taaagaaaaa gaatgggctc 60
gatggaggaa taagaagcta gttatagtca tcggtagaat tgtgaaaggc gcaatttgat 120
tggttaaaat tgttctttga cgagccaacc aattagaaag gaaataagggt gaaggctatt 180
ttacatgtat gcgtcactga cacattgccc aatcagagct ggatattttg aattctttat 240
ttgcatgaaa ggcctataaa aggagagact ctagacacga gctttttattt aagtgcgttc 300
attctcactg ctgttattgt tttctgacag catgcctgaa ccagctaagt cagctcctgc 360
tccgaagaag ggttccaaga aggetgtgac caaggcgcag aagaaggatg gcaagaagcg 420
caagcgcagt cgtaaggaga gctactccgt gtatgtgtac aagggtgctaa aacaggttca 480
ccccgatact ggcattcat ccaaggccat gggcatcatg aattccttcg ttaacgacat 540
cttcgaacgc atcgcaggcg aggettcccc tctggcccac tacaacaagc gctcgacat 600
tacctccagg gagatccaga ccgccgtgcg tctgctgctt cccggagagc tggccaagca 660
cgcagtgtcc gaaggtagca aggetgtcac caagtataca agctccaagt aaatgtgtgc 720
ttaggtgctt taaaactcaa aggetctttt cagagccact caagtctcac ataaagagct 780
ttaatattga atttcaccgt tttctagggg ataagggaat ttttcgattt tgtaatccca 840
gcacttt 847

```

<210> 27
 <211> 2808

-28-

<212> DNA

<213> Homo Sapiens

<400> 27

cgccatgaga ggccagcctg ccagggaaat ccaggaatct gcaacaaaaa cgatgacagt	60
ctgaaataact ctctggtgcc aacctccaaa ttctcgtctg tcaactcaga cccccactag	120
ttgacagagc agcagaatat caactccagt agacttgaat gtgcctcttg gcaaagaagc	180
agagctaacg aggaaagggg tttaaagagt ttttcttggg tgtttgtcaa acttttattc	240
cctgtctgtg tgcagagggg attcaacttc aattttctgc agtggctctg ggtccagccc	300
cttacttaaa gatctggaaa gcatgaagac tgggcctttt ttctatgtc tcttgggaac	360
tgcagctgca atcccgacaa atgcaagatt attatctgat cattccaaac caactgctga	420
aacggtagca cctgacaaca ctgcaatccc cagtttatgg gctgaagctg aagaaaatga	480
aaaagaaaca gcagtatcca cagaagacga ttcccacat aaggctgaaa aatcatcagt	540
actaaagtca aaagaggaaa gccatgaaca gtcagcagaa cagggaaga gttctagcca	600
agagctggga ttgaaggatc aagaggacag tgatgggtcac ttaagtgtga atttgagta	660
tgcaccaact gaaggtacat tggacataaa agaagatatg attgagcctc aggagaaaaa	720
actctcagag aacactgatt ttttggctcc tgggtgtagt tccttcacag attctaacca	780
acaagaaagt atcacaaaga gagaggaaaa ccaagaacaa ctagaaatt attcacatca	840
tcagttgaac aggagcagta aacatagcca aggcctaagg gatcaaggaa accaagagca	900
ggatccaaat atttccaatg gagaagagga agaagaaaaa gagccaggtg aagttggtac	960
ccacaatgat aaccaagaaa gaaagacaga attgcccagg gagcatgcta acagcaagca	1020
ggaggaagac aatacccaat ctgatgatat tttggaagag tctgatcaac caactcaagt	1080
aagcaagatg caggaggatg aatttgatca gggtaaccaa gaacaagaag ataactccaa	1140
tgcagaaatg gaagaggaaa atgcatcgaa cgtcaataag cacattcaag aaactgaatg	1200
gcagagtcaa gagggtaaaa ctggcctaga agctatcagc aaccacaaag agacagaaga	1260
aaagactgtt tctgaggctc tgctcatgga acctactgat gatggtaata ccacgccag	1320
aatcatgga gttgatgatg atggcgatga tgatggcgat gatggcggca ctgatggccc	1380
caggcacagt gcaagtgatg actacttcat cccaagccag gcctttcttg aggccgagag	1440
agctcaatcc attgcctatc acctcaaaat tgaggagcaa agagaaaaag tacatgaaaa	1500
tgaaaatata ggtaccactg agcctggaga gcaccaagag gccaaagaaag cagagaactc	1560
atcaaatgag gaggaaacgt caagtgaagg caacatgagg gtgcatgctg tggattcttg	1620
catgagcttc cagtgtaaaa gaggccacat ctgtaaggca gaccaacagg gaaaacctca	1680

-29-

```

ctgtgtctgc caggatccag tgacttgtcc tccaacaaaa ccccttgatc aagtttgtgg 1740
cactgacaat cagacctatg ctagttcctg tcatctattc gctactaaat gcagactgga 1800
ggggaccaaa aagggggcatc aactccagct ggattatttt ggagcctgca aatctattcc 1860
tacttgtacg gactttgaag tgattcagtt tcctctacgg atgagagact ggctcaagaa 1920
tatcctcatg cagctttatg aagccaactc tgaacatgct ggttatctaa atgagaagca 1980
gagaaataaa gtcaagaaaa tttaacctga tgaaaagagg cttttggctg gggaccatcc 2040
cattgatctt ctcttaaggg actttaagaa aaactaccac atgtatgtgt atcctgtgca 2100
ctggcagttt agtgaacttg accaacaccc tatggataga gtcttgacac attctgaact 2160
tgctcctctg cgagcatctc tggtgcccat ggaacactgc ataaccggtt tctttgagga 2220
gtgtgacccc aacaaggata agcacatcac cctgaaggag tggggccact gctttggaat 2280
taaagaagag gacatagatg aaaatctctt gttttgaacg aagattttta agaactcaac 2340
tttcagcat cctcctctgt tctaaccact tcagaaatat atgcagctgt gatacttgta 2400
gatttatatt tagcaaaatg ttagcatgta tgacaagaca atgagagtaa ttgcttgaca 2460
acaacctatg caccaggtat ttaacattaa ctttggaac aaaaatgtac aattaagtaa 2520
agtcaacata tgcaaaatac tgtacattgt gaacagaagt ttaattcata gtaatttcac 2580
tctctgcatt gacttatgag ataattaatg attaaactat taatgataaa aataatgcat 2640
ttgtattgtt cataatatca tgtgcacttc aagaaaatgg aatgctactc ttttgtgggt 2700
tacgtgtatt attttcaata tcttaatacc ctaataaaga gtccataaaa atccaaaaaa 2760
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaa 2808

```

<210> 28
<211> 2220
<212> DNA
<213> Homo Sapiens

```

<400> 28
ggaaaattac ccggtatcgt tagagctaca ccaaattgc attgagccaa acttgccacc 60
aagagcccaa caatcaccat gatgctgagc acggaaggca gggaggggtt cgtggtgaag 120
gtcaggggcc taccctggtc ctgctcagcc gatgaagtga tgcgcttctt ctctgattgc 180
aagatccaaa atggcacatc aggtattcgt tcatctaca ccagagaagg cagaccaagt 240
ggtgaagcat ttgttgaact tgaatctgaa gaggaagtga aattggcttt gaagaaggac 300
agagaaacca tgggacacag atacgttgaa gtattcaagt ctaacagtgt tgaaatggat 360
tgggtgttga agcatacagg tccgaatagc cctgatactg ccaacgatgg ctctgtccgg 420
cttagaggac tcccatttgg ctgtagcaag gaagagattg ttcagttctt ttcagggttg 480

```

-30-

gaaattgtgc	caaattgggat	gacactgcc	gtggactttc	aggggcgaag	cacaggggaa	540
gcctttgtgc	agtttgett	acaggagata	gctgagaagg	ccttaaagaa	acacaaggaa	600
agaatagggc	acaggtacat	tgagatcttc	aagagtagcc	gagctgaagt	tcgaaccac	660
tatgatcccc	ctcgaaagct	catggctatg	cagcggccag	gtccctatga	taggccgggg	720
gctggcagag	ggtataatag	cattggcaga	ggagctgggt	ttgaaaggat	gaggcgtggt	780
gcctatgggtg	gaggggtatg	aggctatgat	gactatgggtg	gctataatga	tggatatggc	840
tttgggtctg	atagatttgg	aagagacctc	aattactgtt	tttcaggaat	gtctgatcat	900
agatacggag	atgggtgggtc	cagtttccag	agcaccacag	ggcactgtgt	acacatgagg	960
gggttacctt	acagagccac	tgagaatgat	atttataatt	tcttctcacc	tcttaatccc	1020
atgagagtac	atattgaaat	tggacccgat	ggcagagtta	ccggtgaggc	agatgttgaa	1080
tttgctactc	atgaagatgc	tgtggcagct	atggcaaaag	acaaagctaa	tatgcaacac	1140
agatatgtgg	agctcttctt	aaattctact	gcaggaacaa	gtgggggtgc	ttacgatcac	1200
agctatgtag	aacttttttt	gaattctaca	gcaggggcaa	gtggtggcgc	ttatggtagc	1260
caaatgatgg	gagggatggg	cttatccaac	cagtctagtt	atggaggtcc	tgctagccag	1320
cagctgagtg	gtggttatgg	aggtggttat	ggtggtcaga	gcagtatgag	tggatatgac	1380
caagttctgc	aggaaaactc	cagtgactat	cagtcaaacc	ttgcttaggt	agagaaggag	1440
cactaaatag	ctactccaga	tataaaagct	gtacatttgt	gggagttgaa	tagaatggga	1500
gggatgttta	gtatatccag	tatgattggg	aaatgggaaa	tataattgat	tctgatcact	1560
cttggtcagc	ttctctttct	ttatctttct	gtctcctttt	ttaagaaaac	gagttaagtt	1620
taacagtttt	gcattacagg	cttgtgattc	atgcttactg	taaagtggaa	gttgagatta	1680
ttttaaaact	tcaagctcag	taattttgaa	ccactgaaac	attcatctag	gacataataa	1740
caaagttcag	tattgaccat	aactgttaaa	acaattttta	gctttcctca	agttagttat	1800
gtttagaggag	tgtacctaa	cagtaagcgt	atttaggtta	atgcagtttc	acttatgtta	1860
aatgttgctc	ttataccaca	aatacattga	aaacttcgga	tgcagtgtga	gaaacatgcc	1920
tttctgtaaa	actcaaatat	aggagctgtg	tctacgattc	aaagtgaaaa	catttggcac	1980
gtttgttaat	tctagctttt	tggtttaata	tcctgtaagg	cacgtgagtg	tacacttttt	2040
tttttttaaa	ggatacggga	caattttaag	atgtaatacc	aatacttttag	aagtttggtc	2100
gtgtcgtttg	tatgaaaatc	tgaggctttg	gtttaaatct	ttccttgtag	tgtgatttcc	2160
atttagatgt	attgtactaa	gtgaaacttg	ttaaataaat	cttcctttta	aaaactggaa	2220

-31-

<210> 29
 <211> 2203
 <212> DNA
 <213> Homo Sapiens

<400> 29
 cggcgggccgc gccctgggtg ggtccccact gctctcgggg gcgccatgga cgaggccgtg 60
 ggcgacctga agcaggcgct tcctgtgtg gccgagtcgc caacgggtcca cgtggagggtg 120
 catcagcgcg gcagcagcac tgcaaagaaa gaagacataa acctgagtgt tagaaagcta 180
 ctcaacagac ataatatgt gtttgggtgat tacacatgga ctgagtttga tgaacctttt 240
 ttgaccagaa atgtgcagtc tgtgtctatt attgacacag aattaaaggt taaagactca 300
 cagcccacgc atttgagtgc atgcactgtt gcacttcaca ttttcagct gaatgaagat 360
 ggccccagca gtgaaaatct ggaggaagag acagaaaaca taattgcagc aaatcactgg 420
 gttctacctg cagctgaatt ccatgggctt tgggacagct tggatatacga tgtggaagtc 480
 aaatcccatc tcctcgatta tgtgatgaca actttactgt tttcagacaa gaacgtcaac 540
 agcaacctca tcacctggaa ccgggtgggtg ctgctccacg gtctcctgg cactggaaaa 600
 acatccctgt gtaaagcgtt agcccagaaa ttgacaatta gactttcaag caggtaccga 660
 tatggccaat taattgaaat aaacagccac agcctctttt ctaagtgggtt ttcggaaagt 720
 ggcaagctgg taaccaagat gtttcagaag attcaggatt tgattgatga taaagacgcc 780
 ctggtgttcg tgctgattga tgaggtggag agtctcacag ccgcccgaat tgcctgcagg 840
 gcgggcaccg agccatcaga tgccatccgc gtgggtcaatg ctgtcttgac ccaaattgat 900
 cagattaaaa ggcatccaa tgttgtgatt ctgaccactt ctaacatcac cgagaagatc 960
 gacgtggcct tcgtggacag ggctgacatc aagcagtaca ttggggccacc ctctgcagca 1020
 gccatcttca aaatctacct ctcttgtttg gaagaactga tgaagtgtca gatcatatac 1080
 cctcgccagc agctgctgac cctccgagag ctagagatga ttggcttcat tgaaaacaac 1140
 gtgtcaaaat tgagccttct tttgaatgac atttcaagga agagcgaggg cctcagcggc 1200
 cgggtcctga gaaaactccc ctttctgggt catgcgctgt atgtccaggc cccaccgtc 1260
 accatagagg ggttcctcca ggccctgtct ctggcagtg acaagcagtt tgaagagaga 1320
 aagaagcttg cagcttacat ctgatcctgg gcttccccat ctggtgcttt tcccatggag 1380
 aacacacaac cagtaagtga ggttgcccca cacagccgtc tcccaggga tcccttctgc 1440
 aaaccaaacy ttacttagac tgcaagctag aaagccacca aggccaggct ttgttaaaag 1500
 aagtgtattc tatttatgtt gttttaaaat gcatactgag agacaaacat cttgtcattt 1560
 tcactgtttg taaaagataa ttcagattgt ttgtctcctt gtgaagaacc atcgaaacct 1620

-32-

gtttgttccc agcccccccc cagtggatgg gatgcataat gccagcaagt tttgtttaac	1680
agcaaaaaag gaagattaat gcaggtgtta tagaagccag aagagaaact gtgtcaccct	1740
aaagaagcat ataatcatag cattaataat gcacacatta ctccaggtgg aaggtggcaa	1800
ttgctttctg atatcagctc gtttgattta gtgcaaaaat gttttcaaga ctatttaatz	1860
gatgtaaaaa agcctatttc tacattatac caactgagaa aaaaatgggc ggtaaagtgt	1920
tctttcataa taaataatca agacatgggc ccatttgag gaâaagtga gactctgagt	1980
gttccaggga aacacatgct ggacatccct tgtaaccgg tatggggccc cctgcattgc	2040
tgggatgttt ctgcccacgg ttttgtttgt gcaataacgt tatcacattt ctaatgagga	2100
ttcacattaa tataatataa aataaatagg tcagttactg gtctctttct gccgaatgtt	2160
atgttttgct tttatctcac agtaaaataa atataattaa aaa	2203

<210> 30
 <211> 2155
 <212> DNA
 <213> Homo Sapiens

<400> 30	
gtcacatggg gtgcgcgccc agactccgac ccggaggcgg aaccggcagt gcagcccgaa	60
gccccgcagt ccccgagcac gcgtggccat gcgtccctcg cgcgcccgcg ccgcgtgct	120
ggcgctcctg gcctcgctcc tggccgcgcc ccgggtggcc ccggccgagg ccccgcacct	180
ggtgcaggtg gacgcggccc gcgcgtgtg gccctgcgg cgcttctgga ggagcacagg	240
cttctgcccc ccgctgccac acagccaggc tgaccagtac gtcctcagct gggaccagca	300
gctcaacctc gcctatgtgg gcgcgcgtccc tcaccgcggc atcaagcagg tccggaccca	360
ctggctgctg gagcttgtca ccaccagggg gtccactgga cggggcctga gctacaactt	420
caccacctg gacgggtact tggaccttct cagggagaac cagctcctcc cagggtttga	480
gctgatgggc agcgctcgg gccacttcac tgactttgag gacaagcagc aggtgtttga	540
gtggaaggac ttggtctcca gcctggccag gagatacatc ggtaggtacg gactggcgca	600
tgtttccaag tggaacttcg agacgtggaa tgagccagac caccacgact ttgacaacgt	660
ctccatgacc atgcaaggct tcctgaacta ctacgatgcc tgctcggagg gtctgcgcgc	720
cgccagcccc gccctgcggc tgggagggcc cggcgactcc ttccacaccc caccgcgatc	780
cccgctgagc tggggcctcc tgcgccactg ccacgacggt accaacttct tactggggga	840
ggcggggcgtg cggctggact acatctccct ccacaggaag ggtgcgcgca gctccatctc	900
catcctggag caggagaagg tcgtcgcgca gcagatccgg cagctcttcc ccaagttcgc	960
ggacaccccc atttacaacg acgaggcgga cccgctgggtg ggctgggtccc tgccacagcc	1020

-33-

gtggagggcg gacgtgacct acgcggccat ggtggtgaag gtcacgcgc agcatcagaa 1080
 cctgctactg gccaacacca cctccgcctt cccctacgcg ctctgagca acgacaatgc 1140
 cttctgagc taccacccgc accccttcgc gcagcgcacg ctcaccgcgc gcttccaggt 1200
 caacaacacc cgcccgcgc acgtgcagct gttgcgcaag ccggtgctca cggccatggg 1260
 gctgctggcg ctgctggatg aggagcagct ctgggccgaa gtgtcgcagg ccgggaccgt 1320
 cctggacagc aaccacacgg tgggcgtcct ggccagcgcc caccgcccc agggcccggc 1380
 cgacgcctgg cgcgccgcgg tgctgatcta cgcgagcgac gacaccgcgc cccaccccaa 1440
 ccgcagcgtc gcggtgacct tgccgctgcg cggggtgccc cccggccggg gcctggtcta 1500
 cgtcacgcgc tacctggaca acgggctctg cagccccgac ggcgagtggc ggcgcctggg 1560
 ccggcccgctc ttccccacgg cagagcagtt ccggcgcatg cgcgcggctg aggaccgggt 1620
 ggccgcggcg ccccgccctt taccgcgcgg cggccgcctg accctgcgc cgcgcctgcg 1680
 gctgcgcgtc cttttgctgg tgcacgtgtg tgcgcgccc gagaagccgc ccgggcaggt 1740
 cacgcggctc cgcgcctgc ccctgaccca agggcagctg gttctggtct ggtcggatga 1800
 acacgtgggc tccaagtgcc tgtggacata cgagatccag ttctctcagg acggtaaggc 1860
 gtacaccccg gtcagcagga agccatcgac cttcaacctc tttgtgttca gccacagac 1920
 aggtgctgtc tctggctcct accgagttcg agccctggac tactgggccc gaccaggccc 1980
 cttctcggac cctgtgccgt acctggaggt ccctgtgcca agagggcccc catccccggg 2040
 caatccatga gcctgtgctg agccccagtg ggttgacact ccaccggcag tcagcgagct 2100
 ggggctgcac tgtgcccatg ctgccctccc atcacccctt ttgcaatata ttttt 2155

<210> 31
 <211> 7260
 <212> DNA
 <213> Homo Sapiens

<400> 31
 tcaactgtcac tgctaaattc agagcagatt agagcctgcg caatggaata aagtcctcaa 60
 aattgaaatg tgacattgct ctcaacatct cccatctctc tggatttcct tttgcttcat 120
 tattcctgct aaccaattca ttttcagact ttgtacttca gaagcaatgg gaaaaatcag 180
 cagtcttcca acccaattat ttaagtgtcg cttttgtgat ttcttgaagg tgaagatgca 240
 caccatgtcc tcctcgcac tctttctacct ggcgctgtgc ctgctcacct tcaccagctc 300
 tgccacggct ggaccggaga cgctctgcgg ggctgagctg gtggatgctc ttcagttcgt 360
 gtgtggagac aggggctttt atttcaacaa gccacaggg tatggctcca gcagtcggag 420

-34-

ggcgccctcag acagggcatcg tggatgagtg ctgcttccgg agctgtgatc taaggaggct 480
 ggagatgtat tgcgcacccc tcaagcctgc caagtcagct cgctctgtcc gtgcccagcg 540
 ccacaccgac atgcccaga cccagaagga agtacatttg aagaacgcaa gtagagggag 600
 tgcaggaaac aagaactaca ggatgtagga agaccctcct gaggagtga gagtgacatg 660
 ccaccgcagg atccttttgc ctgcacgagt tacctgttaa actttggaac acctacaaa 720
 aaataagttt gataacattt aaaagatggg cgtttccccc aatgaaatac acaagtaaac 780
 attccaacat tgtcttttagg agtgatttgc accttgcaaa aatggctctg gagttggtag 840
 attgctgttg atctttttatc aataatgttc tatagaaaag aaaaaaaaaat atatatatat 900
 atatatttta gtccctgcct ctcaagagcc acaaatgcat ggggtgttgta tagatccagt 960
 tgcactaaat tcctctctga atcttggctg ctggagccat tcattcagca accttgtcta 1020
 agtggtttat gaattgtttc cttatttgca cttctttcta cacaactcgg gctgtttgtt 1080
 ttacagtgtc tgataatctt gttagtctat acccaccacc tcccttcata acctttatat 1140
 ttgccgaatt tggcctcctc aaaagcagca gcaagtcgtc aagaagcaca ccaattctaa 1200
 cccacaagat tccatctgtg gcatttgtac caaatataag ttggatgcat tttatttttag 1260
 acacaaagct ttatttttcc acatcatgct tacaaaaaag aataatgcaa atagttgcaa 1320
 ctttgaggcc aatcattttt aggcataatgt tttaaacata gaaagtttct tcaactcaaa 1380
 agagttcctt caaatgatga gttaatgtgc aacctaatta gtaactttcc tctttttatt 1440
 ttttccatat agagcactat gtaaatttag catatcaatt atacaggata tatcaaacag 1500
 tatgtaaaac tctgtttttt agtataatgg tgctattttg tagtttggtta tatgaaagag 1560
 tctggccaaa acggttaatac gtgaaagcaa aacaataggg gaagcctgga gccaaagatg 1620
 acacaagggg aagggtactg aaaacacccat ccatttggga aagaaggcaa agtcccccca 1680
 gttatgcctt ccaagaggaa cttcagacac aaaagtccac tgatgcaa at tggactggcg 1740
 agtcagaga ggaaactgtg gaatggaaaa agcagaaggc taggaatttt agcagtcctg 1800
 gtttcttttt ctcatggaag aaatgaacat ctgccagctg tgtcatggac tcaccactgt 1860
 gtgaccttg gcaagtcact tcacctctct gtgcctcagt ttctcatct gcaaaatggg 1920
 ggcaatatgt catctaccta cctcaaaggg gtggtataag gtttaaaaag ataaagattc 1980
 agattttttt accctgggtt gctgtaaggg tgcaacatca gggcgcttga gttgctgaga 2040
 tgcaaggaat tctataaata accattcat agcatagcta gagattgggtg aattgaatgc 2100
 tcctgacatc tcagttcttg tcagtgaagc tatccaaata actggccaac tagttgttaa 2160
 aagctaacag ctcaatctct taaaacactt ttcaaaatat gtgggaagca tttgattttc 2220

-35-

aatttgat	ttt tgaattctgc	atttgg	tttt atgaatacaa	agataagtga	aaagagagaa	2280
aggaaaagaa	aaaggagaaa	aacaaagaga	tttctaccag	tgaaagggga	attaattact	2340
ctttgttagc	actcactgac	tcttctatgc	agttactaca	tatctagtaa	aaccttg	2400
aatactataa	ataatattct	attcattttg	aaaaacacaa	tgattccttc	ttttctaggc	2460
aatataagga	aagtgatcca	aaatttgaaa	tattaaaata	atatctaata	aaaagtcaca	2520
aagttatctt	ctttaacaaa	ctttactctt	attcttagct	gtatatacat	ttttttaaaa	2580
agtttg	gttaa aat	atgcttg act	tagagttt	cagttgaaag	gcaaaaactt	2640
ccatcacaac	2640					
aagaaatttc	ccatgcctgc	tcagaaggg	agcccctagc	tctctgtgaa	tgtgttttat	2700
ccattcaact	gaaaattgg	atcaagaaag	tccactgg	agtgtactag	tccatcatag	2760
cctagaaaat	gatccctatc	tgacagatcaa	gattttctca	ttagaacaat	gaattatcca	2820
gcattcagat	ctttctagtc	accttagaac	tttttg	gtta aa	agtaccca	2880
ggcttgatta	2880					
tttcatgcaa	attctatatt	ttacattctt	ggaaagtcta	tatgaaaaac	aaaaataaca	2940
tcttcagttt	ttctccact	gggtcacctc	aaggatcaga	ggccaggaaa	aaaaaaaaag	3000
actccctgga	tctctgaata	tatgcaaaaa	gaaggcccca	tttagtggag	ccagcaatcc	3060
tgttcag	tca aca	agtat	ttt ta	acttga	gcacctcaag	3120
catgcttagc	aatgttctaa	tcactatgga	cagatgtaaa	agaaactata	catcattttt	3180
gccctctgcc	tgttttccag	acatacaggt	tctgtggaat	aagatactgg	actcctcttc	3240
ccaagatggc	acttcttttt	atttcttg	tc cca	gtgt ac	acttcttctc	3300
ctcaacaaaa	ctttataggc	agtcttctgc	agacttaaca	tgttttctgt	catagttaga	3360
tg	tgataatt	ctaagagtgt	ctatgactta	tttcttcac	ttaattctat	3420
ccacagtcaa	3420					
aaatcccca	aggaggaaag	ctgaaagatg	caactgccaa	tattatcttt	cttaactttt	3480
tccaacacat	aatcctctcc	aactggatta	taaataaatt	gaaaataact	cattatacca	3540
attcactatt	ttatttttta	atgaattaaa	actagaaaac	aaattgatgc	aaaccctgga	3600
agtcagttga	ttactatata	ctacagcaga	atgactcaga	tttcatagaa	aggagcaacc	3660
aaaatgtcac	aacccaaaact	ttacaagctt	tgcttcagaa	ttagattgct	ttataattct	3720
tgaatgaggc	aatttcaaga	tatttg	ttaa aga	acagta	acattggtaa	3780
gaatgagctt	3780					
tcaactcata	ggcttatttc	caatttaatt	gaccatactg	gatacttagg	tcaaatttct	3840
gttctctctt	gccc	aaataa	tattaaagta	ttatttgaac	tttttaagat	3900
gaggcagttc	3900					
ccctgaaaaa	gttaatgcag	ctctccatca	gaatccactc	ttctagggat	atgaaaatct	3960
cttaacaccc	accctacata	cacagacaca	cacacacaca	cacacacaca	cacacacaca	4020

-36-

cacacattca ccctaaggat ccaatggaat actgaaaaga aatcacttcc ttgaaaattt	4080
tattaaaaaa caaacaaca aacaaaaagc ctgtccaccc ttgagaatcc ttcctctcct	4140
tggaacgtca atgttttgtt agatgaaacc atctcatgct ctgtggctcc agggtttctg	4200
ttactatttt atgcacttgg gagaaggctt agaataaaag atgtagcaca ttttgctttc	4260
ccatttattg tttggccagc tatgccaatg tgggtgctatt gtttctttaa gaaagtactt	4320
gactaaaaaa aaaagaaaaa agaaaaaaa agaaagcata gacatatattt tttaaagtat	4380
aaaaacaaca attctataga tagatggctt aataaaatag cattaggtct atctagccac	4440
caccaccttt caacttttta tcactcacia gtatgttact gttcaccaaa ttgtgaattt	4500
gggggtgcag gggcaggagt tggaaatttt ttaaagttag aaggctccat tgttttgttg	4560
gctctcaaac ttagcaaaat tagcaatata ttatccaatc ttctgaactt gatcaagagc	4620
atggagaata aacgcgggaa aaaagatctt ataggcaaat agaagaattt aaaagataag	4680
taagttcctt attgattttt gtgcactctg ctctaaaaca gatattcagc aagtggagaa	4740
aataagaaca aagagaaaaa atacatagat ttacctgcaa aaaatagctt ctgccaaatc	4800
ccccttgggt attccttggc atttactggg ttatagaaga cattctccct tcaccagac	4860
atctcaaaga gcagtagctc tcatgaaaag caatcactga tctcatttgg gaaatgttgg	4920
aaagtatttc cttatgagat gggggttctc tactgataaa gaaagaattt atgagaaatt	4980
gttgaaagag atggctaaca atctgtgaag attttttgtt tcttggtttt gttttttttt	5040
ttttttttac tttatacagt ctttatgaat ttcttaatgt tcaaaatgac ttggttcttt	5100
tcttcttttt tttatatcag aatgaggaat aataagttaa acccacatag actctttaaa	5160
actataggct agatagaaat gtatgtttga cttgttgaag ctataatcag actatttaaa	5220
atgttttgct atttttaatc ttaaaagatt gtgctaattt attagagcag aacctgtttg	5280
gctctcctca gaagaaagaa tctttccatt caaatcacat ggctttccac caatattttc	5340
aaaagataaa tctgatttat gcaatggcat catttatattt aaaacagaag aattgtgaaa	5400
gtttatgcc ctcccttgca aagaccataa agtccagatc tggtaggggg gcaacaacaa	5460
aaggaaaatg ttgttgattc ttggtttttg attttgtttt gttttcaatg ctagtgttta	5520
atcctgtagt acatatttgc ttattgctat tttaatatat tataagacct tctgttagg	5580
tattagaaag tgatacatag atatcttttt tgtgtaattt ctattttaaa aagagagaag	5640
actgtcagaa gctttaagtg catatggtac aggataaaga tatcaattta aataaccaat	5700
tcctatctgg aacaatgctt ttgtttttta aagaaacctc tcacagataa gacagaggcc	5760
caggggattt ttgaagctgt ctttattctg ccccatccc aaccagccc ttattatttt	5820

-37-

```

agtatctgcc tcagaatddd atagaggggt gaccaagctg aaactctaga attaaaggaa 5880
cctcactgaa aacatatatt tcacgtgttc cctctctddd ttttctddd tgtgagatgg 5940
ggtctcgcac tgtcccccag gctggagtg agtgatga tctcggtca ctgcaacctc 6000
cacctcctgg gtttaagcga ttctctgccc tcagcctcct gagtagctgg gattacaggg 6060
accaccact atgcccgggt aattdtttgg attdttaata gagacgggggt tttaccatgt 6120
tgccaggtt ggactcaaac tctgacctt gtgatttggc cgcctcagcc tcccaaattg 6180
ctgggattac aggcattgag caccacaccc tgcccatgtg ttccctctta atgtatgatt 6240
acatggatct taaacattat cttctctccc tcattcttca actatctttg atgggggtctt 6300
tcaaggggaa aaaaatccaa gcttdtttta agtaaaaaaa aaaaagaga ggacacaaaa 6360
ccaaatgtta ctgctcaact gaaatattgag ttaagattga gacagagttt ctctaataa 6420
cgggagctga attacctttc actttcaaaa acattgacct ccacaatcct tagaatctgc 6480
cttdttttat attactgagg cctaaaagta aacattctc attdttttt gcccaaatg 6540
cattgatgta aagtaggaaa aataaaaaa gagctctaaa atccctttca agccacctat 6600
tgacccact caccaactca tagcaaagtc acttctgtta atcccttaat ctgattttgt 6660
ttggatattt atcttgtacc cgctgctaaa cacactgcag gagggactct gaaacctcaa 6720
gctgtctact tacatctddd atctgtgtct gtgtatcatg aaaatgtcta ttcaaatat 6780
caaaaccttt caaatattac gcagcttata ttcatgttac ataaaggccc caaatacct 6840
gtcagatctt tttggtaaaa gagttaatga actatgagaa ttgggattac atcatgtatt 6900
ttgcctcatg tatttdttat acacttatag gccaagtgtg ataaataaac ttacagacac 6960
tgaattaatt tcccctgcta ctttgaaacc agaaaataat gactggccat tcgttacatc 7020
tgtcttagtt gaaaagcata ttdtttatta aattaattct gattgtattt gaaattatta 7080
ttcaattcac ttatggcaga ggaatatcaa tctaattgac ttctaaaaat gtaactaatt 7140
gaatcattat cttacattta ctgtttaata agcatatddd gaaaatgtat ggctagagtg 7200
tcataataaa atggtatatc ttdctttagt aattacaaaa aaaaaaaaaa aaaaaaaaaa 7260

```

<210> 32
 <211> 5767
 <212> DNA
 <213> Homo Sapiens

```

<400> 32
gaggaggag agttcacttt tacttcagtg tcagcgcgcg gcggccgtgg ctggctctgg 60
cgagagagca ccgaggaggt gggctgcaga tcttcggggc gctaggggaa atcggcgaga 120
ggcgggatcc gagcgcgcgc gcggggcgca gagccgcga gcctggccag cgagggtagc 180

```

-38-

cgcggggggc	gcgccccggg	cgggcccccg	gagacgcgca	ggatgccaca	cgaagagctg	240
ccgtcgctgc	agagaccccc	ctatggctct	attgtggacg	atgaaaggct	ctctgcagag	300
gagatggatg	agaggaggcg	gcagaacatt	gcttatgaat	atctgtgcca	cttagaggaa	360
gccaaaaggt	ggatggaagt	ttgcttagtt	gaagaattgc	caccaaccac	tgaattggaa	420
gaagggctcc	ggaatggagt	ttaccttgca	aagttagcca	agttctttgc	cccgaaaatg	480
gtatcagaga	aaaagatcta	tgatgtggaa	caaacacgtt	ataagaagtc	tggccttcat	540
tttcgacaca	cagataatac	cgtccagtgg	ttaagagcga	tggagtctat	tggctctacc	600
aagatatatt	atccagaaac	aacagatgtc	tatgatcgga	aaaacatacc	aagaatgata	660
tattgcattc	acgcactgag	tttgtatctg	ttcaaactag	gaatagcacc	ccagatccag	720
gatttgttgg	gcaaagtaga	cttcacagag	gaggaaatca	gtaatatgag	aaaagaactt	780
gagaaatatg	gaatacagat	gccatctttc	agcaaaatag	gtggtattct	ggccaatgaa	840
ctgtccgtgg	atgaagctgc	attacatgct	gcagttatag	ccattaatga	agcagttgaa	900
aaaggaatag	cagagcaaac	cgttgtaaca	ctaagaaacc	caaatgcggt	tttaacttta	960
gtggatgaca	accttgcacc	agaatatcag	aaagaactct	gggatgccaa	aaagaaaaaa	1020
gaggaaaatg	caagactgaa	gaatagctgt	atttcagaag	aagaaagaga	tgcttatgaa	1080
gaactgctga	cacaagcaga	aatccaaggc	aatattaata	aagtcaacag	gcaggctgca	1140
gtggaccata	tcaatgctgt	cattccggaa	ggtgaccccc	agaatacgtc	gcttgcactg	1200
aagaaaccag	aggcccagct	gcctgctggt	tatccctttg	ctgctgccat	gtatcagaac	1260
gaacttttca	acctccagaa	acagaacacc	atgaactact	tggcccacga	ggagcttttg	1320
attgctgtgg	aaatgttgct	tgctgttgct	ttactaaacc	aggccttgga	aagcaacgat	1380
cttgctgtctg	tgacagaatca	actcagaagc	cccgcataag	gcttaaacaa	tctggacaag	1440
gcataatgtg	aacgttatgc	aaacacacta	ctctctgtta	aactagaagt	tttatcccaa	1500
gggcaagata	acttaagctg	gaatgaaatt	cagaattgta	ttgatatggt	taatgctcaa	1560
attcaagaag	aaaatgaccg	agttgtagct	gtagggatca	tcaatgaagc	tattgatgaa	1620
gggaatcctt	tgaggacttt	agaaactttg	ctcctacctc	ctgcgaatat	tagtgatgtg	1680
gaccagccc	atgccagca	ctaccaggat	gttttataacc	atgctaaatc	acagaaactc	1740
ggagactctg	agagtgtttc	caaagtgtt	tggctggatg	agatacagca	agccgtcgat	1800
gaggccaacg	tggacgagga	cagagcaaaa	caatgggtta	ctctgggtgt	tgatgttaat	1860
cagtgtttgg	aaggaaaaaa	atcaagtgat	attttgtctg	tattgaagtc	ttccacttct	1920
aatgcaaatg	acataatccc	ggagtgtgct	gacaaatact	atgatgccct	tgtgaaggca	1980

aaagagctca aatctgaaag agtgtctagt gacggttcat ggctcaaact caacctgcac	2040
aaaaaatatg actactatta caacactgat tcaaaagaga gttcctgggt cacacctgaa	2100
tcatgcttct ataaagaatc atggctcaca ggaaaagaaa tcgaggacat tattgaggaa	2160
gtcacagtag gttacattcg tgagaatata tggctctgctt cagaagagtt gcttcttcgc	2220
tttcaagcca caagctcagg acccatcctt agggaagagt ttgaagctag aaaatcattt	2280
ttgcatgaac aagaagagaa tgtgggtcaaa atacaggctt tttggaaagg atataaacia	2340
cggaaggagt atatgcacag gcggcaaacg ttcattgata atactgattc tgttgtgaag	2400
attcagtcct ggttccgaat ggcaactgca agaaagagct atctttcaag actacagtat	2460
ttcagagatc ataataatga aattgtgaaa atacagtcac tgttgagagc gaacaaagct	2520
agagatgact acaaaacatt ggttggctct gaaaaccac cattaacagt aattcgcaaa	2580
tttgtatacc tgctggacca aagtgatttg gatttccagg aggaactaga ggttgcacga	2640
ttaaggaag aagtagtgac caagatcagg gccaatcaac agctggaaaa agacctgaac	2700
ctgatggaca tcaagattgg actgctggtg aagaacagga tcacactaga ggatgtaatt	2760
tcacacagta aaaagctgaa caagaaaaaa ggaggagaaa tggaaatact gaataacacc	2820
gacaaccaag gaataaaaag tttgagtaag gagaggagaa aaacactaga aacatatcag	2880
cagctgtttt accttttaca gaccaaccct ttataacttg ctaagctgat tttccagatg	2940
ccacagaaca agtccactaa atttatggat actgttattt tcacactata taattatgcc	3000
tctaatacgc gagaagaata tctacttctc aagcttttta aaactgctct ggaggaagaa	3060
ataaaatcaa aagtggacca ggtacaggac atagttactg gtaaccctac agtcatcaag	3120
atggtcgtca gcttcaatag aggtgcccgg ggacagaaca ccctgcgcca actcctggct	3180
ccagtggtaa aagagatcat cgacgacaag tcgctgatta tcaacacaaa ccctgtagag	3240
gtgtacaagg cttgggtgaa ccaactagaa acacagactg gagaggccag caagttgcct	3300
tatgatgtga ccacagaaca agctctaaca taccagaag tgaaaaataa actggaggct	3360
tccattgaga acctgagaag ggtcaccgac aaagtccga attctatcat ttcttccctt	3420
gatctactgc cttatggatt gaggtatata gccaaagtac tgaagaattc gatccatgag	3480
aaattccccg atgcaacaga agatgagcta ttaaagattg ttggaaacct cctgtactat	3540
cggtacatga atccagccat tgtagctcca gatggctttg atatcatcga catgacagct	3600
ggaggtcaga taaattctga ccaaaggaga aacttaggat cagtggccaa ggttcttcag	3660
cacgcagcct ccaacaagct gtttgaagga gaaaatgagc atctctcatc tatgaacaat	3720
tatttatcag agacgtatca ggaattcagg aaatatttca aagaagcatg taatgtccct	3780

-40-

gagccagaag agaagtttaa tatggacaaa tacacagacc tggtagacagt cagcaaacca	3840
gtcatttata tttcaattga agaaatcatc agcacacact cactcctggt ggaacaccag	3900
gatgcaattg cccctgagaa aaatgactta ctgagtgaat tgctgggggtc gctgggagag	3960
gtgccaaaccg tggaatcttt tcttggggaa ggagcagttg accccaatga ccctaacaag	4020
gcaaatacac taagtcagct ttcaaagacc gagatttctc ttgtcttgac aagcaaatat	4080
gacatagagg acggtgaagc tatagatagc cgaagcctca tgataaagac caagaagctg	4140
ataattgatg tgatccggaa ccagccaggg aacacattga cagaaatctt agagacacca	4200
gcaactgcgc aacaggagggt agaccatgcc acggacatgg tgagccgtgc aatgatagat	4260
tccaggactc cagaagaaat gaagcatagc caatctatga ttgaagatgc acagctgcct	4320
cttgagcaga agaagaggaa aatccagagg aatcttcgga cgttggaaaca gactggacac	4380
gtgtcatccg aaaataaata ccaagacatt ctcaatgaga ttgccaaagga tattcgaaat	4440
caaagaatct atcgtaagct tcgaaaagct gaattggcaa aacttcagca gaccctgaat	4500
gcacttaaca agaaggcagc attttatgaa gagcaaatca attattatga cacctacata	4560
aagacttggt tagacaactt aaaaagaaaa aatactcgga gatcaattaa actagatgga	4620
aaaggagaac ccaaaggggc gaagagagcg aagccagtga agtacactgc agcaaagctg	4680
catgagaaag gtgtcctgct agatatagat gatcttcaaa caaaccagtt taagaatggt	4740
acatttgata tcatagctac tgaagatgta ggcattttcg atgtaagatc aaaattcctt	4800
ggtgttgaga tggaaaaggt gcaactcaat attcaggatt tacttcagat gcaatatgaa	4860
ggagtagctg taatgaaaat gtttgataag gttaaagtga atgtaaacct tctcatatac	4920
ctgctgaaca agaagttcta tggaaagtga agtgcctaca gaaatttctt ggattctgta	4980
tcctctggat taggaaatga atttgtttaa tatttttggt tttaaaccatg attgaaatca	5040
ctgcttataa atgtgtgatt ttttttaa at gacccaaact gttctgaaga atgtaccag	5100
gtgccttttt gctaatttga tactataata gaatgagaca taaaatgaat taatggaaac	5160
atatccacac tgtactgtga tatagggtact ctgatttaaa actttggaca tcctgtgatc	5220
tgttttaaag ttgggggggtg ggaaatttag ctgactaggg acaaacatgt aaacctat	5280
tcctatgaaa aaagttttaa atgtccact tgaataacgt aattcttcat agttttttta	5340
atctatggat aaatggaaac ctaattat	5400
gttcttctgg gagttatcaa ttttaaagag aacttttgtg caattcaa at gaagttttta	5460
taagtaattg aaaatgacaa cacaataaca ctttctgtat aaaagtatat attttatgtg	5520
atttattcct actaaatgaa agtgactac tgctcatgt aaagactctt gcacgcagag	5580

-41-

cctttaagtg actaaggaac aacatagata gtgagcatag tccccacctc caccctcac 5640
aatttatattg aatacttcaa ttgtgcctct caattttttg taatgctaaa aaatcagtat 5700
ctagatgggtt tttaaagtga ttctctggaa attgtttttat gtaaaataaa tgttacttaa 5760
ttccatt 5767

<210> 33
<211> 634
<212> DNA
<213> Homo Sapiens

<400> 33
cggctgagag gcagcgaact catctttgcc agtacaggag cttgtgccgt ggcccacagc 60
ccacagccca cagccatggg ctgggacctg acggtgaaga tgctggcggg caacgaattc 120
caggtgtccc tgagcagctc catgtcgggtg tcagagctga aggcgcagat caccagaag 180
attggcgtgc acgccttcca gcagcgtctg gctgtccacc cgagcgggtg ggcgctgcag 240
gacagggtcc cccttgccag ccagggcctg ggccttgcca gcacggctct gctgggtggtg 300
gacaaatgcg acgaacctct gagcatcctg gtgaggaata acaagggccg cagcagcacc 360
tacgaggtcc ggctgacgca gaccgtggcc cacctgaagc agcaagtgag cgggctggag 420
gggtgtgcagg acgacctgtt ctggctgacc ttcgagggga agcccttgga ggaccagctc 480
ccgctggggg agtacggcct caagcccctg agcacctgtg tcatgaatct gcgcctgcgg 540
ggaggcggca cagagcctgg cgggcggagc taagggcctc caccagcatc cgagcaggat 600
caagggccgg aaataaaggc tgttgtaaga gaat 634

<210> 34
<211> 4855
<212> DNA
<213> Homo Sapiens

<400> 34
gaattcccct cccccctttt tccatgcagc tgatctaaaa gggaataaaa ggctgcgcat 60
aatcataata ataaaagaag gggagcgcga gagaaggaaa gaaagccggg aggtggaaga 120
ggaggggggag cgtctcaaag aagcgatcag aataataaaa ggaggccggg ctctttgcct 180
tctggaacgg gccgctcttg aaagggtttt tgaaaagtgg tgttgttttc cagtcgtgca 240
tgctccaatc ggcggagtat attagagccg ggacgcggcc gcaggggcag cggcgacggc 300
agcaccggcg gcagcaccag cgcgaacagc agcggcggcg tcccagtgct ccgcggcggc 360
gcgcgcagcg atgcgttccc cacggacacg cggccgggtc gggcgccccc taagcctcct 420
gctcgccctg ctctgtgccc tgcgagccaa ggtgtgtggg gcctcgggtc agttcgagtt 480

ggagatcctg tccatgcaga acgtgaacgg ggagctgcag aacgggaact gctgcggcgg	540
cgcccggaac ccgggagacc gcaagtgcac ccgcgacgag tgtgacacat acttcaaagt	600
gtgcctcaag gagtatcagt cccgcgtcac ggccgggggg ccctgcagct tcggctcagg	660
gtccacgcct gtcacggggg gcaacacctt caacctcaag gccagccgcg gcaacgaccc	720
gaaccgcac gtgctgcctt tcagtttcgc ctggccgagg tcctatacgt tgcttggtga	780
ggcgtgggat tccagtaatg acaccgttca acctgacagt attattgaaa aggcttctca	840
ctcgggcatg atcaacccca gccggcagtg gcagacgctg aagcagaaca cgggcgttgc	900
ccactttgag tatcagatcc gcgtgacctg tgatgactac tactatggct ttggctgtaa	960
taagttctgc cgcccagag atgacttctt tggacactat gcctgtgacc agaattggca	1020
caaaacttgc atggaaggct ggatggggcc cgaatgtaac agagctatct gccgacaagg	1080
ctgcagtcct aagcatgggt cttgcaaact ccaggtgac tgcaggtgcc agtacggctg	1140
gcaaggcctg tactgtgata agtgcacccc acaccggga tgcgtccacg gcactctgaa	1200
tgagccctgg cagtgcctct gtgagaccaa ctggggcggc cagctctgtg acaaagatct	1260
caattactgt gggactcatc agccgtgtct caacggggga acttgtagca acacaggccc	1320
tgacaaatat cagtgttctt gccctgaggg gtattcagga cccaactgtg aaattgctga	1380
gcacgcctgc ctctctgac cctgtcacia cagaggcagc tgtaaggaga cctccctggg	1440
ctttgagtgt gagtggtccc caggctggac cggcccaca tgctctacaa acattgatga	1500
ctgttctcct aataactgtt cccacggggg cacctgccag gacctggtta acggatttaa	1560
gtgtgtgtgc cccccacagt ggactgggaa aacgtgccag ttagatgcaa atgaatgtga	1620
ggccaaacct tgtgtaaacg ccaaactctg taagaatctc attgccagct actactgca	1680
ctgtcttccc ggctggatgg gtcagaattg tgacataaat attaattgact gccttgcca	1740
gtgtcagaat gacgcctcct gtcgggattt ggttaatggg tategctgta tctgtccacc	1800
tggctatgca ggcgatcact gtgagagaga catcgatgaa tgtgccagca acccctgttt	1860
gaatgggggt cactgtcaga atgaaatcaa cagattccag tgtctgtgtc ccactggttt	1920
ctctggaaac ctctgtcagc tggacatcga ttattgtgag cctaatacct gccagaacgg	1980
tgcccagtgc tacaaccgtg ccagtgacta tttctgcaag tgccccgagg actatgaggg	2040
caagaactgc tcacacctga aagaccactg ccgcacgacc ccctgtgaag tgattgacag	2100
ctgcacagtg gccatggctt ccaacgacac acctgaaggg gtgcgggtata tttcctccaa	2160
cgtctgtggg cctcacggga agtgcaagag tcagtcggga ggcaaattca cctgtgactg	2220
taacaaaggc ttcacgggaa catactgcca tgaaaatatt aatgactgtg agagcaaccc	2280

ttgtagaaac ggtggcactt gcatcgatgg tgtcaactcc tacaagtgca tctgtagtga	2340
cggctgggag ggggcctact gtgaaaccaa tattaatgac tgcagccaga acccctgcc	2400
caatgggggc acgtgtcgcg acctgggtcaa tgacttctac tgtgactgta aaaatgggtg	2460
gaaaggaaag acctgccact cacgtgacag tcagtgtgat gaggccacgt gcaacaacgg	2520
tggcacctgc tatgatgagg gggatgcttt taagtgcattg tgcctggcg gctgggaagg	2580
aacaacctgt aacatagccc gaaacagtag ctgcctgcc aaccctgcc ataatggggg	2640
cacatgtgtg gtcaacggcg agtcctttac gtgcgtctgc aaggaaggct gggaggggcc	2700
catctgtgct cagaatacca atgactgcag ccctcatccc tgttacaaca gcggcacctg	2760
tgtggatgga gacaactggg accggtgcga atgtgccccg ggttttgctg ggcccgactg	2820
cagaataaac atcaatgaat gccagtcttc acctgtgccc tttggagcga cctgtgtgga	2880
tgagatcaat ggctaccggg gtgtctgccc tccagggcac agtggtgcca agtgccagga	2940
agtttcaggg agaccttgca tcaccatggg gagtgtgata ccagatgggg ccaaaggga	3000
tgatgactgt aatacctgcc agtgcctgaa tggacggatc gcctgctcaa aggtctggtg	3060
tggccctcga ccttgctgc tccacaaagg gcacagcgag tgccccagcg ggcagagctg	3120
catccccatc ctggacgacc agtgcttcgt ccaccctgc actggtgtgg gcgagtgtcg	3180
gtcttcaggt ctccagccgg tgaagacaaa gtgcacctct gactcctatt accaggataa	3240
ctgtgcgaac atcacattta cttttaacaa ggagatgatg tcaccaggtc ttactacgga	3300
gcacatttgc agtgaattga ggaatttgaa tattttgaag aatgtttccg ctgaatattc	3360
aatctacatc gcttgcgagc cttcccttc agcgaacaat gaaatacatg tggccatttc	3420
tgctgaagat atacgggatg atgggaaccc gatcaaggaa atcactgaca aaataatcga	3480
tcttgtagt aaacgtgatg gaaacagctc gctgattgct gccgttgag aagtaagagt	3540
tcagaggcgg cctctgaaga acagaacaga tttccttggt cccttgctga gctctgtctt	3600
aactgtggct tggatctggt gcttgggtgac ggcttctac tgggtgctgc ggaagcggcg	3660
gaagccgggc agccacacac actcagcttc tgaggacaac accaccaaca acgtgcggga	3720
gcagctgaac cagatcaaaa acccattga gaaacatggg gccaacacgg tccccatcaa	3780
ggattacgag aacaagaact ccaaaatgtc taaaataagg acacacaatt ctgaagtaga	3840
agaggacgac atggacaaac accagcagaa agcccgggtt gccaaagcagc cggcgtacac	3900
gctggtagac agagaagaga agcccccaa cggcagccg acaaaacacc caaactggac	3960
aaacaaacag gacaacagag acttggaag tgcccagagc ttaaaccgaa tggagtacat	4020
cgtatagcag accgcgggca ctgccggcg taggtagagt ctgagggtt gtagttcttt	4080

-44-

```

aaactgtcgt gtcatactcg agtctgaggc cgttgctgac ttagaatccc tgtgttaatt 4140
tagtttgaca agctggctta cactggcaat ggtagttctg tggttggctg ggaaatcgag 4200
tggcgcatct cacagctatg caaaaagcta gtcaacagta cccctggttg tgtgtcccct 4260
tgcagccgac acggtctcgg atcaggtctc caggagctgc ccagccccct ggtactttga 4320
gctcccactt ctgccagatg tctaattggtg atgcagtctt agatcatagt tttattttata 4380
tttattgact cttgagttgt ttttgtatat tggttttatg atgacgtaca agtagttctg 4440
tatttgaaag tgcctttgca gctcagaacc acagcaacga tcacaaatga ctttattatt 4500
tatttttttt aattgtattt ttgttggttg gggaggggag actttgatgt cagcagttgc 4560
tggtaaaatg aagaatttaa agaaaaatg tccaaaagta gaactttgta tagttatgta 4620
aataattctt ttttattaat cactgtgtat atttgattta ttaacttaat aatcaagagc 4680
cttaaaacat cattcctttt tatttatatg tatgtgttta gaattgaagg tttttgatag 4740
cattgtaagc gtatggcttt atttttttga actcttctca ttacttggtg cctataagcc 4800
aaaaaggaaa ggggtgtttg aaaatagttt attttaaaac aataggatgg gctac 4855

```

```

<210> 35
<211> 9534
<212> DNA
<213> Homo Sapiens

```

```

<400> 35
cagcgactcc tctggctccc gagaagtgga tccggtcgcg gccactacga tgccgggagc 60
cgccgggggc ctctccttc tgctgctctc cggaggcctc gggggcgtag aggcgcagcg 120
gccgcagcag cagcggcagt cacaggcaca tcagcaaaga ggtttattcc ctgctgtcct 180
gaatcttgct tctaattgct ttatcagcag caatgcaaca tgtggagaaa aaggacctga 240
aatgtactgc aaattggtag aacatgtccc tgggcagcct gtgaggaacc cgcagtgtcg 300
aatctgcaat caaaacagca gcaatccaaa ccagagacac ccgattacaa atgctattga 360
tggaagaac acttggtggc agagtcccag tattaagaat ggaatcgaat accattatgt 420
gacaattaca ctggatttac agcaggtggt ccagatcgcg tatgtgattg tgaaggcagc 480
taactcccc cggcctggaa actggatttt ggaacgctct cttgatgatg ttgaatacaa 540
gccctggcag tatcatgctg tgacagacac ggagtgccta acgctttaca atatttatcc 600
ccgcactggg ccaccgtcat atgccaaaga tgatgaggtc atctgcactt cattttactc 660
caagatacac cccttagaaa atggagagat tcacatctct ttaatcaatg ggagaccaag 720
tgccgatgat ccttctccag aactgctaga atttacctcc gctcgtata ttcgcctgag 780

```

-45-

atttcagagg atccgcacac tgaatgctga cttgatgatg tttgctcaca aagacccaag	840
agaaattgac cccattgtca ccagaagata ttactactcg gtcaaggata tttcagttgg	900
agggatgtgc atctgctatg gtcatgccag ggcttgtcca cttgatccag cgacaaataa	960
atctcgctgt gagtgtgagc ataacacatg tggcgatagc tgtgatcagt gctgtccagg	1020
attccatcag aaaccctgga gagctggaac ttttctaact aaaactgaat gtgaagcatg	1080
caattgtcat ggaaaagctg aagaatgcta ttatgatgaa aatgttgcc aagaatct	1140
gagtttgaat atacgtggaa agtacattgg aggggggtgc tgcattaatt gtacccaaaa	1200
cactgctggt ataaactgag agacatgtac agatggcttc ttcagacca aaggggtatc	1260
tccaaattat ccaaggccat gccagccatg tcattgcatg ccaattggtt ccttaaata	1320
agtctgtgtc aaggatgaga aacatgctcg acgaggtttg gcacctggat cctgtcattg	1380
caaaactggt tttggaggtg tgagctgtga tcggtgtgcc aggggctaca ctggctaccc	1440
ggactgcaaa gcctgtaact gcagtgggtt agggagcaaa aatgaggatc cttgttttgg	1500
ccctgtatc tgcaaggaaa atgttgaagg aggagactgt agtcgttgca aatccggctt	1560
cttcaatttg caagaggata attggaaagg ctgcgatgag tgtttctgtt caggggtttc	1620
aaacagatgt cagagttcct actggaccta tggcaaaata caagatatga gtggctggta	1680
tctgactgac cttcctggcc gcattcgagt ggctccccag caggacgact tggactcacc	1740
tcagcagatc agcatcagta acgcggaggc ccggcaagcc ctgccgcaca gctactactg	1800
gagcgcgccg gctccctatc tgggaaacaa actcccagca gtaggaggac agttgacatt	1860
taccatatca tatgaccttg aagaagagga agaagataca gaacgtgttc tccagcttat	1920
gattatctta gagggtaatg acttgagcat cagcacagcc caagatgagg tgtacctgca	1980
cccatctgaa gaacatacta atgtattgtt acttaaagaa gaatcattta ccatacatgg	2040
cacacatttt ccagtccgta gaaaggaatt tatgacagtg cttgcgaatt tgaagagagt	2100
cctcctacaa atcacataca gctttgggat ggatgccatc ttcaggttga gctctgttaa	2160
ccttgaatcc gctgtctcct atcctactga tgggaagcatt gcagcagctg tagaagtgtg	2220
tcagtgccca ccagggtata ctggctcctc ttgtgaatct tgttggccta ggcacaggcg	2280
agttaacggc actatTTTTT gtggcatctg tgagccatgt cagtgctttg gtcatgcgga	2340
gtcctgtgat gacgtcactg gagaatgcct gaactgtaag gatcacacag gtggcccata	2400
ttgtgataaa tgtcttctcg gtttctatgg cgagcctact aaaggaacct ctgaagactg	2460
tcaaccctgt gcctgtccac tcaatatccc atccaataac tttagcccaa cgtgccattt	2520
agaccggagt cttggattga tctgtgatgg atgccctgtc gggtagacag gaccacgtg	2580

-46-

tgagaggtgt gcagaaggct attttggaca accctctgta cctggaggat catgtcagcc	2640
atgccaatgc aatgacaacc ttgacttctc catccctggc agctgtgaca gcttgtctgg	2700
ctcctgtctg atatgtaaac caggtacaac aggccggtac tgtgagctct gtgctgatgg	2760
atatttttga gatgcagttg atgcgaagaa ctgtcagccc tgtcgctgta atgccggtgg	2820
ctctttctct gaggtttgcc acagtcaaac tggacagtgt gagtgcagag ccaacgttca	2880
gggtcagaga tgtgacaaat gcaaggctgg gacctttggc ctacaatcag caaggggctg	2940
tgttcctctg aactgcaatt cttttgggtc taagtcattc gactgtgaag agagtggaca	3000
atgttggtgc caacctggag tcacagggaa gaaatgtgac cgctgtgccc acggctattt	3060
caacttccaa gaaggaggct gcacagcttg tgaatgttct catctgggta ataattgtga	3120
cccaaagact gggcgatgca tttgcccacc caataccatt ggagagaaat gttctaaatg	3180
tgcacccaat acctggggcc acagcattac cactggttgt aaggcttgta actgcagcac	3240
agtgggatcc ttggatttcc aatgcaatgt aaatacaggc caatgcaact gtcattccaaa	3300
attctctggt gcaaaatgta cagagtgcag tcgaggtcac tggaactacc ctcgctgcaa	3360
tctctgtgac tgcttctctc ctgggacaga tgccacaacc tgtgattcag agactaaaaa	3420
atgctcctgt agtgatcaaa ctgggcagtg cacttgtaag gtgaatgtgg aaggcatcca	3480
ctgtgacaga tgccggcctg gcaaattcgg actcgatgcc aagaatccac ttggctgcag	3540
cagctgctat tgcttcggca ctactaccca gtgctctgaa gcaaaaggac tgatccggac	3600
gtgggtgact ctgaaggctg agcagaccat tctaccctg gtagatgagg ctctgcagca	3660
cacgaccacc aagggcattg tttttcaaca tccagagatt gttgccaca tggacctgat	3720
gagagaagat ctccatttgg aaccttttta ttggaaactt ccagaacaat ttgaaggaaa	3780
gaagttgatg gcctatgggg gcaaactcaa gtatgcaatc tatttcgagg ctcgggaaga	3840
aacaggtttc tctacatata atcctcaagt gatcattcga ggtgggacac ctactcatgc	3900
tagaattatc gtcaggcata tggtgctcc tctgattggc caattgacaa ggcatgaaat	3960
tgaaatgaca gagaaagaat ggaaatatta tggggatgat cctcgagtcc atagaactgt	4020
gacccgagaa gacttcttgg atatactata tgatattcat tacattctta tcaaagctac	4080
ttatggaaat ttcatgcgac aaagcaggat ttctgaaatc tcaatggagg tagctgaaca	4140
aggacgtgga acaacaatga ctccctccagc tgacttgatt gaaaaatgtg attgtcccct	4200
gggctattct ggcctgtcct gtgaggcatg cttgccggga ttttatcgac tgcgttctca	4260
accagggtggc cgcacccctg gaccaaccct gggcacctgt gttccatgtc aatgtaatgg	4320
acacagcagc ctgtgtgacc ctgaaacatc gatatgccag aattgtcaac atcacactgc	4380

-47-

tggtgacttc	tgtgaacgat	gtgctcttgg	atactatgga	attgtcaagg	gattgccaaa	4440
tgactgtcag	caatgtgcct	gccctctgat	ttcttccagt	aacaatttca	gcccctcttg	4500
tgtcgcagaa	ggacttgacg	actaccgctg	cacggcttgt	ccacggggat	atgaaggcca	4560
gtactgtgaa	aggtgtgccc	ctggctatac	tggcagtgca	ggcaaccctg	gaggctcctg	4620
ccaagaatgt	gagtgtgatc	cctatggctc	actgcctgtg	ccctgtgacc	ctgtcacagg	4680
attctgcacg	tgccgacctg	gagccacggg	aaggaagtgt	gacggctgca	agcactggca	4740
tgcacgcgag	ggctgggagt	gtgttttttg	tggagatgag	tgcactggcc	ttcttctcgg	4800
tgacttggct	cgctgggagc	agatgggtcat	gagcatcaac	ctcactggtc	cgctgcctgc	4860
gccatataaa	atgctgtatg	gtcttgaaaa	tatgactcag	gagctaaagc	acttgctgtc	4920
acctcagcgg	gccccagaga	ggcttattca	gctggcagag	ggcaatctga	atacactcgt	4980
gaccgaaatg	aacgagctgc	tgaccagggc	taccaaagtg	acagcagatg	gcgagcagac	5040
cggacaggat	gctgagagga	ccaacacaag	agcaaagtc	ctgggagaat	tcattaagga	5100
gcttgcccgg	gatgcagaag	ctgtaaatga	aaaagctata	aaactaaatg	aaactctagg	5160
aactcgagac	gaggcctttg	agagaaat	ggaagggctt	cagaaagaga	ttgaccagat	5220
gattaaagaa	ctgaggagga	aaaatctaga	gacacaaaag	gaaattgctg	aagatgagtt	5280
ggtagctgca	gaagcccttc	tgaaaaaagt	gaagaagctg	tttggagagt	cccgggggga	5340
aaatgaagaa	atggagaagg	atctccggga	aaaactggct	gactacaaaa	acaaagttag	5400
tgatgcttgg	gaccttttga	gagaagccac	agataaaatc	agagaagcta	atcgcttatt	5460
tgacgataat	cagaaaaaca	tgactgcatt	ggagaaaaag	aaggaggctg	ttgagagcgg	5520
caaacgacaa	attgagaaca	ctttaaaga	aggcaatgac	atactcgatg	aagccaaccg	5580
tcttgcatg	gaaatcaact	ccatcataga	ctatgttgaa	gacatccaaa	ctaaattgcc	5640
acctatgtct	gaggagctta	atgataaaat	agatgacctc	tccaagaaa	taaaggacag	5700
gaagcttgct	gagaaggtgt	cccaggctga	gagccacgca	gctcagttga	atgactcatc	5760
tgctgtcctt	gatggaatcc	ttgatgaggc	taaaaacatc	tccttcaatg	ccactgcagc	5820
cttcaaagct	tacagcaata	ttaaggacta	tattgatgaa	gctgagaaag	ttgccaaaga	5880
agccaaagat	cttgcatatg	aagctacaaa	actggcaaca	ggtcctcggg	gtttattaaa	5940
ggaagatgcc	aaaggctgtc	ttcagaaaag	cttcaggatt	cttaacgaag	ccaagaagtt	6000
agcaaatgat	gtaaaagaaa	atgaagacca	tctaaatggc	ttaaaaacca	ggatagaaaa	6060
tgctgatgct	agaaatgggg	atctcttgag	aactttgaat	gacacttttg	gaaagttatc	6120
agctattcca	aatgatacag	ctgctaaact	gcaagctgtt	aaggacaaag	ccagacaagc	6180

caacgacaca gctaaagatg tactggcaca gattacagag ctccaccaga acctcgatgg 6240
cctgaagaag aattacaata aactagcaga cagcgctgcc aaaacgaatg ctgtgggttaa 6300
agatccttcc aagaacaaaa tcattgccga tgcagatgcc actgtcaaaa atttagaaca 6360
ggaagctgac eggctaatag ataaactcaa acccatcaag gaacttgagg ataacctaaa 6420
gaaaaacatc tctgagataa aggaattgat aaaccaagct cggaaacaag ccaattctat 6480
caaagtatct gtgtcttcag gaggtgactg cattcgaaca tacaaccag aatcaagaa 6540
aggaagttac aataatattg ttgtcaacgt aaagacagct gttgctgata acctcctctt 6600
ttatcttggg agtgccaaat ttattgactt tctggctata gaaatgcgta aaggcaaagt 6660
cagcttcctc tgggatgttg gatctggagt tggacgtgta gagtaccag atttgactat 6720
tgatgactca tattggtacc gtatcgtagc atcaagaact gggagaaatg gaactatttc 6780
tgtgagagcc ctggatggac ccaaagccag cattgtgcc agcacacacc attcgacgtc 6840
tcctccaggg tacacgattc tagatgtgga tgcaaatgca atgctgtttg ttggtggcct 6900
gactgggaaa ttaaagaagg ctgatgctgt acgtgtgatt acattcactg gctgcatggg 6960
agaaacatac tttgacaaca aacctatagg tttgtggaat ttccgagaaa agaaggtga 7020
ctgcaaagga tgcactgtca gtccctcagg ggaagatagt gaggggacta ttcaatttga 7080
tggagaaggt tatgcattgg tcagccgtcc cattcgctgg taccccaaca tctccactgt 7140
catgttcaag ttcagaacat tttcttcgag tgctcttctg atgtatcttg ccacacgaga 7200
cctgagagat ttcatgagtg tggagctcac tgatgggcac ataaaagtca gttacgatct 7260
gggctcagga atggcttccg ttgtcagcaa tcaaaaccat aatgatggga aatggaaatc 7320
attcactctg tcaagaattc aaaaacaagc caatatatca attgtagata tagatactaa 7380
tcaggaggag aatatagcaa ctctgtcttc tggaaacaac tttggtcttg acttgaaagc 7440
agatgacaaa atatattttg gtggcctgcc aacgctgaga aacttgagta tgaaagcaag 7500
gccagaagta aatctgaaga aatattccgg ctgcctcaaa gatattgaaa tttcaagaac 7560
tccgtacaat atactcagta gtcccgatta tgttggtgtt accaaaggat gttccctgga 7620
gaatgtttac acagttagct ttcctaagcc tggttttgtg gagctctccc ctgtgccaat 7680
tgatgtagga acagaaatca acctgtcatt cagcaccaag aatgagtccg gcatcattct 7740
tttgggaagt ggaggggacac cagcaccacc taggagaaaa cgaaggcaga ctggacaggc 7800
ctattatgta atactcctca acaggggccg tctggaagtg catctctcca caggggcacg 7860
aacaatgagg aaaattgtca tcagaccaga gccgaatctg tttcatgatg gaagagaaca 7920
ttccgttcat gtagagcgaa ctagaggcat ctttacagtt caagtggatg aaaacagaag 7980

-49-

atacatgcaa aacctgacag ttgaacagcc tatcgaagtt aaaaagcttt tcgttggggg	8040
tgctccacct gaatttcaac cttccccact cagaaatatt cctccttttg aaggctgcat	8100
atggaatctt gttattaact ctgtcccat ggactttgca aggctgtgt cttcaaaaa	8160
tgctgacatt ggtcgctgtg cccatcagaa actccgtgaa gatgaagatg gagcagctcc	8220
agctgaaata gttatccagc ctgagccagt tcccaccca gcccttccta cgccaccccc	8280
agttctgaca catggctctt gtgctgcaga atcagaacca gctcttttga tagggagcaa	8340
gcagttcggg ctttcaagaa acagtcacat tgcaattgca tttgatgaca ccaaagttaa	8400
aaaccgtctc acaattgagt tggaagtaag aaccgaagct gaatccggct tgctttttta	8460
catggctgcg atcaatcatg ctgattttgc aacagttcag ctgagaaatg gattgcccta	8520
cttcagctat gacttgggga gtggggacac ccacaccatg atccccacca aaatcaatga	8580
tggccagtgg cacaagatta agataatgag aagtaagcaa gaaggaattc tttatgtaga	8640
tggggcttcc aacagaacca tcagtccaa aaaagccgac atcctggatg tcgtgggaat	8700
gctgtatgtt ggtgggttac ccatcaacta cactaccga agaattggtc cagtgacct	8760
tagcattgat ggctgctca ggaatctcca catggcagag gccctgccc atctggaaca	8820
accacctcc agcttccatg ttgggacatg ttttgcaa atgctcagagg gaacatattt	8880
tgacggaacc ggttttgcca aagcagttgg tggtattcaa gtgggattgg accttcttgt	8940
agaatttgaa ttcgagacaa ctacaacgac tggagttctt ctggggatca gtagtcaaaa	9000
aatggatgga atgggtattg aaatgattga tgaaaagttg atgtttcatg tggacaatgg	9060
tgccggcaga ttcactgctg tctatgatgc tggggttcca gggcatttgt gtgatggaca	9120
atggcataaa gtcactgcca acaagatcaa acaccgcatt gagctcacag tcgatgggaa	9180
ccaggtggaa gcccaaagcc caaaccagc atctacatca gctgacacaa atgaccctgt	9240
gtttgttga ggcttccag atgacctcaa gcagtttggc ctaacaacca gtattccgtt	9300
ccgaggttgc atcagatccc tgaagctcac caaaggcaca gcaagccact ggaggttaat	9360
tttgccaagg ccctggaact gaggggctt caacctgtat catgcccagc caactaataa	9420
aaataagtgt aaccccagga agagtctgtc aaaacaagta tatcaagtaa aacaaacaaa	9480
tatatatttac ctatatatgt taattaaact aatttgtgca tgtacataga attc	9534

<210> 36
 <211> 5683
 <212> DNA
 <213> Homo Sapiens

<400> 36
 ccgcccgtg ttgcgctcct tcccagaatc cgctccggcc ttctcttct gccgcgattc 60

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☒ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)